

Die Sprache in Kinderzeitschriften
Ein korpusbasierter Vergleich der Sprache in *GEolino* und *GEO*

Masterarbeit
zur Erlangung des akademischen Grades
Master of Arts (M.A.)
im Fach Linguistik

Humboldt-Universität zu Berlin
Philosophische Fakultät II
Institut für deutsche Sprache und Linguistik

Eingereicht von Nadine Lestmann
Matrikelnummer: 186815

Erstgutachterin:
Zweitgutachterin:

Prof. Dr. Anke Lüdeling
Prof. Dr. Katharina Spalek

Berlin, den 24.04.2012

Inhaltsverzeichnis

1. Einleitung	5
2. Material und Hypothesen	7
2.1 Definition von Sachtexten	7
2.2 <i>GEOLino</i> und <i>GEO</i> als Sachtexte	7
2.3 Hypothesen	8
3. Theoretischer Überblick	9
3.1 Forschungslücke	9
3.1.1 Vernachlässigung des Aspekts <i>Sprache</i>	9
3.1.2 Kritik an der Vernachlässigung des Aspekts <i>Sprache</i>	11
3.1.3 Exkurs 1: Autoren und ihr junges Publikum	12
3.2 Forschungslage	13
3.2.1 Oberflächenmerkmale: Wort- und Satzlänge	14
3.2.2 Wortbildung	17
3.2.3 Adjektivfrequenz	17
3.2.4 Verhältnis von attributiven zu prädikativen Adjektiven	17
3.2.5 Komparation	18
3.2.6 Abstrakta	18
3.2.7 Konjunktionsfrequenz	19
3.2.8 Semantische Klassen von Konjunktionen	20
3.2.9 Leserkontakt	20
3.3 Psycholinguistische Grundlage: Kognitiver Entwicklungsstand	23
3.4 Exkurs 2: Registerbegriff	24
4. Methode	25
4.1 Korpora: Textauswahl und Variablen	26
4.1.1 Themengleichheit	27
4.1.2 Metadaten	27
4.1.3 Autoren	28
4.1.4 Metasprachliche Elemente	28
4.1.5 Eingriffe	29

4.1.6 Erstellungswerkzeug	29
4.1.7 Zusammensetzung der Korpora	29
4.2 Annotation und Richtlinien	30
4.2.1 Oberflächenmerkmale	30
4.2.2 Wortbildung: Komposition	31
4.2.3 Anzahl der Hauptglieder	32
4.2.4 Komplexität der Hauptglieder	33
4.2.5 Adjektivfrequenz	34
4.2.6 Attributive und prädikative Adjektive	34
4.2.7 Komparation	34
4.2.8 Adjektivische Amplifikation	35
4.2.9 Nominale Abstrakta	36
4.2.10 Konjunktionsfrequenz	37
4.2.11 Semantische Klassen von Konjunktionen	37
4.2.12 Leserkontakt	38
4.2.13 Metasprachliche Elemente	39
5. Analyse	39
5.1 Allgemeine Erwartungen	39
5.2 Hauptkomponentenanalyse	40
5.3 Univariater Vergleich	42
5.3.1 Voranalyse	43
5.3.2 Hauptanalyse	45
5.3.2.1 Oberflächenmerkmale	45
5.3.2.2 Wortbildung: Komposition	46
5.3.2.3 Anzahl der Hauptglieder	46
5.3.2.4 Komplexität der Hauptglieder	46
5.3.2.5 Adjektivfrequenz	47
5.3.2.6 Attributive und prädikative Adjektive	47
5.3.2.7 Komparation	48
5.3.2.8 Adjektivische Amplifikation	48
5.3.2.9 Nominale Abstrakta	49
5.3.2.10 Konjunktionsfrequenz	49

5.3.2.11 Semantische Klassen von Konjunktionen	50
5.3.2.12 Leserkontakt	50
5.3.3 Kritik	51
6. Zusammenfassung.....	51
7. Ausblick.....	54
7.1 Optimierung und Erweiterung der Korpora	54
7.2 Registerfrage	56
7.3 Doppeltexte	56
7.4 Aspekt <i>Verständlichkeit</i>	57
8. Literaturverzeichnis.....	58
A. Richtlinien	62
B. Lexikalisierte Komposita	66
C. Grafiken zum univariaten Vergleich.....	67
D. Doppeltexte: Verteilung der POS-tags	81
9. Selbstständigkeitserklärung	83

1. Einleitung

Lesen ist eine Beschäftigung, die uns durch das ganze Leben begleitet. Teilweise schon vor Schulbeginn von uns gelernt, profitieren wir in den verschiedensten Situationen von der Fähigkeit, geschriebene Texte zu entschlüsseln; ob wir nun aus schulischen, beruflichen oder aus Unterhaltungsgründen zu literarischen Erzeugnissen greifen.

Dabei spielen Sachtexte eine wichtige Rolle in der Wissensvermittlung und zwar schon bei jungen Lesern. Sie werden von den Kindern nicht nur im schulischen Kontext in Form von Schulbüchern gelesen, sondern auch außerhalb der Schule als neues Hobby in Form von Kinderzeitschriften. Hier dienen Sachtexte einer besonderen Art von Wissensvermittlung, die gerade bei Kindern eine zentrale Rolle in der kognitiven Entwicklung spielt und auch in meiner Arbeit von zentraler Bedeutung ist: die Verbindung von Information und Unterhaltung.¹

Im Zusammenhang mit der PISA-Studie betont Berthold (2010), dass sich im Zeitraum von 2000 bis 2007 die Zahl der Kindersachbücher verdoppelt hat. Kinderzeitschriften erfreuen sich in Deutschland einer großen Beliebtheit. Dabei setzt sich die *GEOLINO* seit 2009 als wissensvermittelnde Zeitschrift mit ca. 240.000 verkauften Exemplaren deutlich von den Kinderzeitschriften ab, die einen rein unterhaltenden Anspruch haben und steht noch vor der Micky Maus (ca. 195.000 verkaufte Exemplare), die als Spitzenreiter im Bereich der rein unterhaltenden Zeitschriften angesehen werden kann². Baurmann (2009) berichtet von einer Umfrage unter 10- bis 18jährigen Schülern, die Angaben darüber machen sollten, welche Texte sie für die Schule und in der Freizeit nutzen. Die Umfrage ergab, dass viele der genannten Texte in die Kategorie *Sachtext* fallen. Auch nach Rosebrock/Nix (2008:74) gilt ab der 5. Klasse der Grundsatz „reading to learn“ statt „learning to read“.

Dies zeigt, dass Sachtexte für Kinder eine wichtige Rolle spielen: Sie bieten den Kindern die Möglichkeit, auf unterhaltende Weise Neues zu lernen und sich dieses Wissen auch eigenständig anzueignen, zum Beispiel in Form von Kinderzeitschriften.

Damit Wissensvermittlung durch Sachtexte gelingt, steht der Autor in einer großen Verantwortung: Er muss auf unterhaltsame Art und Weise eine Fülle von Information im Text unterbringen; dem Leser auf ansprechende Art die Welt erklären. Diese Verbindung von erfolgreicher Informationsvermittlung und Unterhaltung ist keineswegs trivial.

Nach Gierlich (2005) und Niederhauser (1999:117) hat der Wissenschaftsjournalist – als der der Autor von Sachtexten angesehen werden kann – die Aufgabe, fachliche Information zu übersetzen. Dabei fungiert er als Experte, der seinem fachexternen Publikum – dem Laien – wissenschaftliche Inhalte erfolgreich vermitteln soll. Dies gelingt ihm durch Popularisierung:

„Populärwissenschaftliche Wissensvermittlung bedeutet *Transformation, Transfer, Umsetzung* oder *Übersetzung* wissenschaftlicher Inhalte in fachexterne Darstellungen unter Anwendung bestimmter Methoden, Techniken und Strategien der Popularisierung.“

¹Zur Geschichte der Kinderpresse vgl. Pech (2010:11ff.), Ulf Diederichs (2010:48ff.), Baacke/Lauffer (1994:175f), Sommer (1994:28ff.).

²http://www.gujmedia.de/_content/20/02/200273/GEOLINO._2009.pdf vom 7.5.2010.

Auch für erwachsene Leser spielen unterhaltende Sachtexte eine wichtige Rolle in der Wissensaneignung und -erweiterung (zum Beispiel durch die Zeitschriften *GEO*, *National Geographic Deutschland* oder *Welt der Wunder*). Aber auch wenn junge Leser und erwachsene Leser hier die gleiche Klasse von Texten lesen, lassen sich anscheinend ganz unterschiedliche Popularisierungs-Strategien (im Sinne Niederhausers) beobachten. Legt man uns nämlich je eine Passage aus einer Kinderzeitschrift und einer Zeitschrift für Erwachsene vor, würden wir wohl intuitiv entscheiden können, welche Passage aus welcher Zeitschrift stammt.

Doch worauf beruht die Intuition? Wie unterscheiden sich Kinder- und Erwachsenenzeitschriften? Gibt es neben den offensichtlichen inhaltlichen und (typo-)grafischen Unterschieden auch linguistische Unterschiede, sprachliche Merkmale also, die unsere Entscheidung steuern?

In meiner Arbeit möchte ich die Frage beantworten, worin sich die in Kinderzeitschriften verwendete Sprache von der Sprache in Erwachsenenzeitschriften unterscheidet. Genauer: Ich betrachte die Zeitschrift *GEOLino*, deren Kernzielgruppe sich aus älteren Grundschulern im Alter von 10 bis 12 Jahren zusammensetzt und möchte die verwendete Sprache möglichst genau beschreiben. Dadurch möchte ich eine linguistische Abgrenzung nach oben hin zu der entsprechenden Erwachsenenzeitschrift *GEO* schaffen. Die zentrale Frage meiner Arbeit lautet also:

(1) Gibt es systematische sprachliche Unterschiede zwischen *GEOLino* und *GEO*?

Wenn es tatsächlich linguistische Unterschiede gibt, kann man einen Schritt weiter gehen und fragen:

(2) Wird in Kinderzeitschriften ein besonderes sprachliches Register verwendet?

In diesem Zusammenhang ergibt sich auch die Frage, was ein spätkindliches Sprachregister ausmacht (dazu mehr in Abschnitt 3.4).

Eine empirische Herangehensweise an die Thematik bietet einen wesentlichen Vorteil gegenüber einer rein theoretischen Arbeit: Durch eine Korpus basierte Studie kann ich die sprachliche Charakteristik realer Texte für junge Leser untersuchen und somit direkt beobachten, welche Sprache diesen Lesern gegenüber tatsächlich verwendet wird. Kramer (1990:65f) bemerkt, dass

„eine linguistische Bestimmung und Einteilung von Sachtexten bisher nicht existiert [...] Ein fruchtbarer Weg, eine Lösung des Problems „Sachtext“ zu finden, ist u.U. der, nicht beim Text in abstracto, sondern beim konkreten Textexemplar anzusetzen und in aufsteigender Weise vorzugehen [...] Dazu müsste eine Teilmenge von Texten desselben Geltungsbereichs zugrunde gelegt werden.“

Diesen Versuch werde ich hier vornehmen. Letztendlich soll diese Arbeit dazu beitragen, die sprachliche Charakteristik von Kinder- und Erwachsenenzeitschriften genauer zu fassen und entsprechend zu formulieren. Dadurch kann meine Studie grundlegend für weitere Forschung auf diesem Gebiet sein (vgl. Abschnitt 7.).

In den folgenden Abschnitten erläutere ich meine Hypothesen und mein Material genauer und gebe einen theoretischen Überblick über Grundlagen und Forschungsliteratur, die für meine Studie relevant sind. Der zweite Teil der Arbeit ist praktisch ausgerichtet. Dort finden meine Erkenntnisse aus dem theoretischen Teil Anwendung in Form der

eigentlichen Korpusarbeit. Die Analysen sollen zeigen, inwiefern sich die theoretisch gewonnenen Erkenntnisse in der Praxis widerspiegeln und an welchen Punkten sich Theorie und Praxis vielleicht widersprechen.

2. Material und Hypothesen

Vor der Hypothesenbildung ist es sinnvoll, zunächst zu klären, was ein Sachtext eigentlich ist. Welche Eigenschaften muss ein Text aufweisen, um in die Kategorie *Sachtext* eingeordnet zu werden? Und weshalb gehören die von mir betrachteten *GEolino*- und *GEO*-Texte in diese Kategorie?

2.1 Definition von Sachtexten

Ähnlich wie Niederhauser sehen Baumann (2009) und Gierlich (2005) Sachtexte allgemein als Teil der fachexternen Kommunikation. Die Kategorie *Sachtext* lässt sich in verschiedenen Hinsichten weiter unterteilen³, die einzelnen Untergruppen sollen hier jedoch nicht weiter ausgeführt werden.

Allen Sachtexten ist gemeinsam, dass sie dem Leser bestimmte Sachverhalte erklären sollen (vgl. Kühn (2010:129)). Sie sind nach Rosebrock/Nix (2008) semantisch domänenspezifisch, handeln also von bestimmten fachlichen Wissensgebieten. Steinlein (2010:32) legt fest, dass Sachtexte unterhalten sollen, jedoch in erster Linie informierende Funktion haben und daher „die informierend-berichtenden Teile als solche dominieren“. Damit heben sie sich von belletristischen Texten ab. Entgegen Hummelberger (2003:332) stelle ich hier die Vermittlung von Wissen als zentrale Aufgabe und wichtigstes Merkmal von Sachtexten in den Vordergrund. Dabei grenze ich sie von rein wissenschaftlichen Texten auf der einen Seite und von belletristischen Texten auf der anderen Seite ab. Ich verstehe Sachtexte in meiner Arbeit als Texte, die

- (1) dem Leser in erster Linie Wissen zu einem bestimmten Thema vermitteln und
- (2) sich nicht an den Experten, sondern an den Laien als Leser richten.

Im Rahmen meiner Arbeit ist diese grobe und natürlich nicht erschöpfende Definition von Sachtexten völlig ausreichend. Sie soll lediglich dazu dienen, die Texte der *GEolino* und *GEO* kategorisch einzuordnen.

Eine feingliedrige Klassifizierung von Sachtexten ist problematisch, da sich in der Forschungsliteratur sehr verschiedene Ansätze der Systematisierung finden (vgl. auch Hummelsberger 2003:331). Deswegen wirkt der Blick auf die Kategorie *Sachtext* zunächst etwas undifferenziert (vgl. auch Meissner (2010)). Eine erschöpfende Kategorisierung ist jedoch nicht das Ziel meiner Arbeit und soll daher hier nicht versucht werden.

2.2 *GEolino* und *GEO* als Sachtexte

Das Kriterium der Unterhaltung sehe ich als Untergruppe der Sachtexte, wie ich sie hier verstehe. Meiner Meinung nach sind unterhaltende Sachtexte als Spezifizierung zu verstehen, wodurch sie sich von Sachtexten in Schulbüchern unterscheiden.

³ Je nach Herangehensweise lassen sich Sachtexte zum Beispiel funktional (vgl. Kühn (2010:129)), nach Autorintention oder inhaltlich unterteilen (vgl. auch Meissner (2010)).

Wie kann man nun die Texte aus der *GEOLino* und *GEO* einordnen? Es liegt auf der Hand, dass sie die beiden von mir genannten Kriterien (1) und (2) erfüllen. Beide Zeitschriften haben das Ziel, einem interessierten Laien Wissen zu vermitteln. Daher lassen sie sich – je nach Autor – der Untergruppe *darstellende*, *informierende* oder *instruierende Sachtexte* mit dem Ziel der Wissensvermittlung zuordnen (vgl. Gierlich (2005:26f)) bzw. der Untergruppe *didaktische* oder *Lehrtexte* mit dem Ziel des Erinnerns bzw. des Aufbaus von deklarativem Wissen (vgl. Christmann/Groebe (2006:150), Rosebrock/Nix(2008:75)).

Sie weisen jedoch noch ein weiteres Merkmal auf, nämlich die schon angesprochene Verbindung von Unterhaltung und Information. Das Verhältnis der beiden Aspekte entspricht Steinleins Definition von Sachtexten: Die Wissensvermittlung steht im Vordergrund.

Texte von diesem Format kann man nach Wittwen (1995:24) als *Infotainment* bezeichnen, da hier Informationen auf unterhaltende Art und Weise für den Leser aufbereitet werden.

Die Kriterien *Mischung von Information und Unterhaltung* sowie *Wissensvermittlung als primäres Ziel* erfüllen die Zeitschriften *GEOLino* und *GEO* gleichermaßen. Sie unterscheiden sich jedoch bezüglich des Parameters *Zielgruppe*: Die Kernzielgruppe der *GEOLino* setzt sich aus älteren Grundschulkindern im Alter von 10 bis 12 Jahren zusammen^{4,5}. Die *GEO* richtet sich hingegen an erwachsene Leser⁶.

Beide Zeitschriften werden sowohl von männlichen als auch von weiblichen Lesern rezipiert (ebd.).

Mit der Einordnung und Erläuterung des Materials ist es nun einfacher, genaue Hypothesen zu formulieren.

2.3 Hypothesen

Aus meiner Fragestellung (1) (siehe Abschnitt 1) lässt sich folgende zentrale Hypothese ableiten, die ich in meiner Arbeit überprüfe:

- (1) Die Sprache in der *GEOLino* unterscheidet sich systematisch von der Sprache in der *GEO*.

Zudem lässt sich noch eine weitere Hypothese formulieren, die direkt mit dem Vergleich zwischen *GEOLino* als Zeitschrift für junge Leser und *GEO* als Mutterblatt zusammenhängt:

- (2) *GEOLino*-Leser gehören zu einer Zwischengruppe zwischen Leseanfängern und Lesern auf Erwachsenen-Niveau.

Deußing (1927) hat Hypothese (2) durch eine Aufsatzstudie mit Schülern aus der 2., 3., 4., 6. und 8. Klasse anhand der Wortlänge (in Silben) bereits bestätigt. Er vergleicht die Texte der Schüler mit literarischen Texten für Erwachsene. Dabei stellt er fest dass die Schüler ab Klasse 4 mit der Zahl der verwendeten Einsilber quantitativ genau zwischen den jüngeren Schülern und den Texten für Erwachsene liegen. Sie stellen ebenfalls in Bezug auf die Mehrsilber eine Zwischengruppe zwischen den jungen Schülern und den Texten für Erwachsene dar (vgl. Deußing (1927:117ff.)).

⁴http://www.gujmedia.de/_content/20/02/200273/GEOLINO._2009.pdf vom 7.5.2010.

⁵<http://ems.guj.de/print/portfolio/geolino/leserschaft/> vom 9.4.2012.

⁶http://ems.guj.de/uploads/tx_hngujmediasales/pdfs/GEO_Profil_2012_Nov_2011.pdf vom 9.4.2012.

Auch bei Gibson/Levin (1989:137f) finden sich Hinweise darauf, dass die sprachliche Beschaffenheit der Texte für ältere Grundschüler (5. Klasse) zwischen den Texten für Leseanfänger und den Texten für Erwachsene liegt. Die relevanten Parameter sind hier die Wortlänge, die Größe des Vokabulars und die Anzahl der Wiederholungen. Hypothese (2) ließe sich anhand der *GEO*-Zeitschriftenreihe gut überprüfen, denn hier liegt die angenommene Dreiteilung des Leseniveaus in Form von drei verschiedenen Zeitschriften bereits vor: *GEOMini* für Leseanfänger, *GEOlino* für ältere Grundschulkinder und *GEO* für erwachsene Leser. Wenn es sprachliche Merkmale in der *GEOlino* gibt, die sich von denen der *GEOMini* und der *GEO* signifikant unterscheiden, sollte es möglich sein, diese *GEOlino*-typischen Merkmale herauszufiltern. Im Rahmen meiner Arbeit kann ich Hypothese (2) jedoch nicht überprüfen, weil ich hier zunächst nur die Abgrenzung nach oben hin versuche.

Um Hypothese (1) zu prüfen, ist zunächst ein umfassender Blick in die Forschungsliteratur nötig.

3. Theoretischer Überblick

In den folgenden Abschnitten erläutere ich, inwiefern es ein Forschungsdefizit hinsichtlich der Sprachbeschreibung in Sachtexten gibt. Ich gehe genauer auf die Arbeiten ein, die Aufschluss darüber geben, wie die Sprache in Sachtexten für junge Leser beschaffen ist. Zudem erkläre ich, in welchem Stadium des Lese- und Spracherwerbs sich die Zielgruppe der *GEOlino* ungefähr befindet und worin sie sich von der Zielgruppe der *GEO* kognitiv unterscheidet. Ein kurzer Exkurs soll schließlich dazu dienen, den in Fragestellung (2) (siehe Abschnitt 1) angesprochenen Begriff *Register* grob zu erläutern.

3.1 Forschungslücke

Die von mir betrachteten Arbeiten zur Sprache in Sachtexten bzw. Sachtexten für junge Leser lassen sich im Wesentlichen in drei Bereiche einteilen:

- (a) Arbeiten, die den Aspekt *Sprache* gar nicht oder nur ungenau betrachten
- (b) Arbeiten, die eine fehlende oder ungenaue Betrachtung des Aspekts *Sprache* kritisieren
- (c) Arbeiten, die genauer auf den Aspekt *Sprache* eingehen

Die Arbeiten aus den Bereichen (a) und (b) zeigen in Bezug auf Hypothese (1) eine Lücke in der Forschungsgrundlage, da die Behandlung der sprachlichen Seite von Sachtexten in diesen Arbeiten durchaus problematisch ist. In den folgenden beiden Abschnitten werde ich zunächst die Arbeiten aus den Bereichen (a) und (b) vorstellen, um die Forschungslücke zu skizzieren. Die Arbeiten aus dem Bereich (c) sollen mir letztendlich sprachliche Merkmale liefern, die ich als Vergleichsgrundlage meiner Studie verwenden kann. Daher erläutere ich diese Arbeiten in Abschnitt 3.2 genauer.

3.1.1 Vernachlässigung des Aspekts *Sprache*

Es gibt zahlreiche Studien über die Sprache zu Kleinkindern und Leseanfängern (*child directed speech*), jedoch nicht zur Zielgruppe der Kinderzeitschriften für die von mir betrachtete Altersklasse. Die Literatur zur Kinderpresse bezieht sich primär auf inhaltliche

Aspekte, die in meiner Arbeit keine Rolle spielen. In einigen wenigen Arbeiten greifen die Autoren zwar den Aspekt *Sprache* auf, dieser ist jedoch durchgängig ungenau klassifiziert und wird lediglich peripher erwähnt. So unterscheiden Baacke/Lauffer (1994:14) die Sprachstile "kindlich-naiv", "kindorientiert", "seriös", "sachlich", "jugendorientiert" und "Fachjargon", formulieren jedoch nur sehr ungenau bzw. gar nicht, was charakteristisch für den jeweiligen Sprachstil ist. In ihrer Diplomarbeit über den Markt der deutschen Kinderzeitschriften beschreibt Kurz, dass in Kinderzeitschriften "Anleitungen [...] ausführlich und verständlich erklärt" (Kurz 2000:44), "Reportagen [...] verständlich, aber gleichzeitig sachlich geschrieben" (Kurz 2000:46) und "Länderberichte [...] umfassend, aber verständlich geschrieben" sind (Kurz 2000:56). Die linguistischen Kriterien der erwähnten Verständlichkeit und die sprachlichen Merkmale allgemein werden jedoch auch hier nicht weiter ausdifferenziert. Auch in anderen Arbeiten zur Kinderpresse in Deutschland habe ich vergleichbar ungenaue Definitionen und Klassifikationen gefunden.

Selbst in einer Arbeit, die sich genau an einem Brennpunkt der Wissensvermittlung durch Sachtexte befindet, wird das Kriterium der sprachlichen Angemessenheit und damit der Aspekt *Sprache* nicht weiter erläutert: Im Rahmen der *Internationalen Grundschul-Lese-Untersuchung* (IGLU) wurde die Lesekompetenz von Viertklässlern im internationalen Vergleich getestet. Dazu mussten die Schüler literarische sowie Informationstexte lesen und Verständnisfragen zu den Texten beantworten. Im Zusammenhang mit der Textauswahl betonen Bos et al. (2003:92), dass „Curriculumsexperten für den Bereich Leseverständnis benannt [wurden], die die Lesetexte [...] nach vorgegebenen Kriterien beurteilen sollten“.

Eines der Kriterien lautet „Angemessenheit von Sprache und Inhalt für Viertklässler“ (ebd.), was jedoch nicht weiter differenziert wird. Die Einschätzungen der Experten bezüglich dieses Kriteriums sind so nicht nachvollziehbar. Bos et al. (2003:72) bemerken lediglich, dass ein Text „spezifische Sprach- und Strukturelemente [beinhaltet], welche die Lesbarkeit beeinflussen“, gehen jedoch nicht weiter darauf ein.

Ebenfalls im didaktischen Bereich erläutern Rosebrock/Nix (2008) die Charakteristik von Sachtexten und wie man diese Art der Lektüre unterstützen kann. Doch werden Lehrtexte "als Textsorte absurder Weise [sic] während des schulischen Curriculums an keiner Stelle systematisch erarbeitet" (Rosebrock/Nix(2008:75)).

Es stellt sich auch die Frage, wonach sich Autoren eigentlich richten, wenn sie Texte für junge Leser verfassen. Journalistenschulen in Deutschland bieten keine Kurse an, in denen man lernt, für Kinder zu schreiben.⁷ Ich gehe in Abschnitt 3.1.3 noch genauer darauf ein. Ebenso bieten die Lesealterempfehlungen der Verlage keinen Anhaltspunkt, da es in diesem Bereich keine öffentlich zugänglichen und verbindlichen Kriterien und Kontrollinstanzen gibt, wie etwa die FSK⁸ der Filmwirtschaft oder die USK⁹ für die Altersfreigabe von Computerspielen. Selbst Rahmenpläne für den Deutschunterricht sind für die Beschreibung des Aspekts *Sprache* wenig aussagekräftig, da auch hier keine

⁷ Eine wichtige Quelle liefert mir ein Gespräch mit der Journalistin Kathrin Kommerell, die u.a. für die Kinderzeitschrift *Benjamin* schreibt und Autorin des Buches *Journalismus für junge Leser* ist.

⁸ Freiwillige Selbstkontrolle

⁹ Unterhaltungssoftware Selbstkontrolle

konkreten sprachlichen Anforderungen an die Schüler festgelegt sind, sondern eher inhaltlich zwischen Leseanfängern und fortgeschrittenen Lesern unterschieden wird.¹⁰

Die sprachliche Seite von Texten wird also nur ungenau bzw. gar nicht betrachtet. Im Allgemeinen geht man mit Texten eher einseitig um¹¹, daher ist das Wissen über die sprachliche Komponente von Sachtexten defizitär. Ich habe keine Arbeit finden können, die diese Komponente sowohl wissenschaftlich fundiert als auch umfangreich erschließt.

3.1.2 Kritik an der Vernachlässigung des Aspekts *Sprache*

Sieber (2008) bemängelt die einseitige Betrachtung von Texten und kritisiert, dass man sich bei der Bewertung von Texten schnell nur noch auf den *Inhaltsproblemraum* konzentriert und den *Sprachproblemraum* nicht weiter betrachtet. Das führt dazu, dass man über die sprachliche Seite von Texten nur sehr schwer diskutieren kann, weil objektive Betrachtungskriterien fehlen. Nach Sieber (2008:272)

"haben wir uns alle auf irgendeine Weise eine Sprache dafür zurechtgelegt – die einen eher mit grammatischen und stilistischen, die anderen mit rhetorischen Kategorien. Aber das, was Texte zu guten Texten macht, bleibt meist ziemlich schummrig".

Moers (2004:8) geht noch einen Schritt weiter und spricht sich gegen eine einseitige Betrachtung von Texten aus, die die kognitive Komponente des Lesers nicht mit einbezieht:

"Ein Text enthält zwar Sprach- und Strukturelemente, jedoch entsteht die Bedeutung des Gelesenen erst durch die Interaktion der Kinder mit dem Text vor dem Hintergrund einer Leseerfahrung."

Sieber und Moers beschreiben hier drei Aspekte, die relevant für eine Bewertung eines Textes sind: sein Inhalt, seine sprachliche Beschaffenheit und textexterne Faktoren, die den Leser, seine Vorerfahrung, Erwartung und kognitiven Fähigkeiten betreffen.

In meiner Arbeit werde ich die Texte nicht bewerten, sondern lediglich ihre linguistische Beschaffenheit zu beschreiben versuchen. Ich möchte die sprachlichen Elemente, aus denen die Texte aufgebaut sind, möglichst genau formulieren.

Wie wichtig die Betrachtung der linguistischen Beschaffenheit von Texten ist, betont auch Antos (2008:243), der von einer Verantwortlichkeit des Autors für Formulierungen spricht, „Denn gerade die Formulierungen eines Textes sind [...] durchaus folgenreich für die Verständnisbildung“. Hier wird ein direkter Einfluss der sprachlichen Mittel auf das Verstehen eines Textes deutlich.

Im *Hamburger Leitfaden für die Bewertung von Textprodukten* (HLBT) beschäftigt sich May (1996) mit der Bewertung von kindlichen Bild- und Textprodukten. Er operationalisiert bestimmte Fähigkeiten von Kindern, die für die Produktion von Bildern

¹⁰http://www.berlin.de/imperia/md/content/sen-bildung/schulorganisation/lehrplaene/gr_deu_1_6.pdf?start&ts=1157977375&file=gr_deu_1_6.pdf vom 17.4.12.

¹¹ Bos et al. (2003:92) fassen die Aspekte *Inhalt* und *Sprache* zu einem Parameter zusammen.

und Texten relevant sind und versucht, „Gesichtspunkte für die Förderung dieser Fähigkeiten im Unterricht zu entwickeln“ (May (1996:3)).

Dazu nennt May verschiedene Merkmale, die als Kriterien für eine Beschreibung und letztendlich Bewertung von kindlichen Bildern und Texten dienen können. Diese Kriterien für den HLBT sind empirisch motiviert: Sie wurden auf der Basis von Texten entwickelt, die ca. 4000 Kinder in den Klassen 1 bis 5 produziert haben.

Auch May (1996:44) führt dabei die Relevanz der sprachlichen Seite von Texten auf, denn er bewertet die Texte unter anderem nach dem Kriterium "Sprachliche Angemessenheit (Wortschatz und Satzkonstruktionen), Sprachstil / Schriftsprachlichkeit".

Christmann/Groeben (2006) beschäftigen sich mit der Kategorisierung von Sachtexten und gehen genauer auf die für meine Arbeit relevanten didaktischen bzw. Lehrtexte ein. Die zentrale Funktion dieser Texte sehen Christmann/Groeben (2006:151) im Aufbau von deklarativem Wissen. Sie stellen in diesem Zusammenhang die Frage, welche textseitigen (im Sinne von sprachlichen) Merkmale den Aufbau einer entsprechenden kognitiven Struktur beim Leser erleichtern. Auch hier wird also eingeräumt, dass die sprachliche Seite eines Textes relevant für das Verständnis des Textes ist.

Kramer (1990:64ff.) beschäftigt sich in ihrer Dissertation mit der Klassifikation von Sachtexten und nennt „sprachlich-kommunikative Merkmale“, die einen Sachtext ausmachen. Diese Merkmale seien jedoch „keine solide Basis für eine Beschreibung der spezifischen sprachlichen Gestaltung von Sachtexten“ (ebd.). Sie kommt zu dem Schluss, dass „es nicht möglich [ist], 'Sachtext' als linguistische Kategorie umfassend und grundlegend zu definieren“ (ebd.).

Ich möchte das Defizit in der Betrachtung der sprachlichen Seite von Sachtexten für junge Leser mit meiner Arbeit verkleinern und letztendlich dazu beitragen, diese Forschungslücke zu schließen.

3.1.3 Exkurs 1: Autoren und ihr junges Publikum

Trotz der großen Verantwortung, die Autoren von Sachtexten ihrem jungen Publikum gegenüber tragen, ist der Kontakt zur Zielgruppe offenbar ein Problem. Nach Kommerell (Quelle: persönliches Gespräch) mangelt es nicht nur an wissenschaftlichen Quellen zu Sprache und Stil in Kinderzeitschriften; es gibt auch kaum eine Zielgruppenforschung.

Sommer (1994) erstellt eine Marktübersicht über regelmäßig erscheinende Kinderzeitschriften und -seiten in Deutschland, indem er umfangreiche Fragebögen an die einzelnen Zeitschriften- und Zeitungsredaktionen verschickt. Die Fragebögen beziehen sich auf verschiedene Bereiche, unter anderem auf die „wirtschaftliche und personelle Situation“ sowie „Konzeptionen und Zielgruppen“ (Sommer (1994:84)). Er stellt hier die zentrale Frage: *Wer schreibt eigentlich für die Kinder?* Sommers Umfrage ergibt, dass die wenigsten Autoren, die regelmäßig für Kinder schreiben, eine entsprechende journalistische Zusatzqualifikation haben, die sie dafür prädestinieren würde. Sie schreiben *aus dem Bauch heraus* und versuchen, den Leser aus der Ferne einzuschätzen bzw. die Lesegewohnheiten der Kinder aus ihrem privaten Umfeld auf ihr Publikum zu übertragen (vgl. Sommer (1994:103)).

Die Autoren sind gezwungen, explorativ vorzugehen, eben weil es keine eindeutigen Anleitungen gibt, wie man für Kinder schreibt. Es gibt eine Fülle von Empfehlungen und

Verboten, die jedoch meist introspektiv begründet sind. An vielen Stellen überschneiden sich diese Intuitionen jedoch. Deswegen muss man sich fragen, ob hier eine Regelmäßigkeit bezüglich bestimmter sprachlicher Merkmale vorliegt. In diesem Fall könnte man eventuell doch Kriterien herausfiltern, die sich operationalisieren lassen. Man müsste entscheiden, ob sich dadurch ein Kriterienraster aufbauen lässt, wonach man sich beim Verfassen von Sachtexten für Kinder richten kann.

3.2 Forschungslage

Es gibt eine ganze Reihe von Arbeiten, die – wenn auch oft implizit und nur vage – über die sprachliche Seite von Sachtexten Aufschluss geben. Dies sind die in Abschnitt 3.1 genannten Arbeiten aus dem Bereich (c): Arbeiten, die genauer auf den Aspekt *Sprache* eingehen. Für sich genommen liefert jede von ihnen nur Erkenntnisse über einen kleinen Teil der sprachlichen Seite; zusammengenommen bieten sie jedoch eine verhältnismäßig große Sammlung von lexikalischen, syntaktischen und pragmatischen Merkmalen, die von den Autoren als (un)charakteristisch bzw. (un)angemessen für junge Leser eingestuft werden.

Ich beziehe die sprachlichen Merkmale für meine Arbeit aus verschiedenen Forschungsrichtungen:

- (1) Textlinguistik/Sachtextforschung
- (2) Lesbarkeitsforschung/Textverständlichkeitsmodelle
- (3) Registerforschung
- (4) Psycholinguistik
- (5) Journalismus

Dabei betrachte ich sowohl wissenschaftlich fundierte Arbeiten als auch Arbeiten, die keine wissenschaftliche Grundlage haben und eher *aus dem Bauch heraus* argumentieren. Weil es in der wissenschaftlichen Fachliteratur nicht viele konkrete Anhaltspunkte für die Sprache in Kinderzeitschriften bzw. Sachtexten für junge Leser gibt, erweitere ich meine Recherche auf den journalistischen Bereich. Hier gibt es oft keinen wissenschaftlichen Anspruch, daher sind die Autoren viel freier und deutlicher in der Formulierung sprachlicher Merkmale.

Im Bereich der Psycholinguistik beschäftige ich mich mit Arbeiten zur Entwicklung der Sprachproduktionsfähigkeiten von Kindern. Aus Erkenntnissen zur Sprachproduktion kann man schließen, welche sprachlichen Phänomene den Kindern bereits vertraut sind. Generell geht man bei der Sprachentwicklung von dem Grundsatz aus, dass das Verstehen der Produktion voraussetzt (vgl. zum Beispiel Clark (2009:283), Dietrich (2007:97), Höhle (2010:128f)). Auf der Grundlage dieser Annahme ziehe ich den Umkehrschluss für das Sprachverstehen: Wenn Doderer (1960:317ff.) beispielsweise zeigt, dass Kinder schon im frühen Grundschulalter einfach eingebettete Nebensätze verwenden, haben sie die Funktionsweise dieser Konstruktion offenbar längst verstanden, denn sie nutzen sie produktiv. Also gehe ich davon aus, dass das Merkmal *einfacher Nebensatz* in der *GEOLino* uneingeschränkt verwendet wird.

Nach Adamzik (2008:166) kann man einen Text nie auf alle grammatischen bzw. lexikalischen Merkmale untersuchen, da es zu viele Merkmale gibt. Daher ist auch in

meiner Arbeit eine Auswahl nötig. Die Sammlung der sprachlichen Merkmale ist nicht nur groß, sondern auch sehr weit gestreut und unausgeglichen, was die Tiefe der Merkmale betrifft. Ich habe insgesamt 10 Hauptmerkmale ausgewählt, die verschiedene sprachliche Bereiche betreffen und in mehreren Arbeiten angesprochen werden.

In den folgenden Abschnitten erläutere ich die Arbeiten genauer, die sich mit dem jeweiligen sprachlichen Merkmal beschäftigen. Diese Arbeiten stellen gleichermaßen die Motivationsgrundlage für die von mir ausgewählten Merkmale dar. Oft werden sprachliche Merkmale lediglich erwähnt und nicht weiter diskutiert. In diesen Fällen nehme ich die Merkmale in meine Untersuchung auf, sofern sie mir relevant erscheinen.

3.2.1 Oberflächenmerkmale: Wort- und Satzlänge

Insbesondere in den Arbeiten, die sich mit Lesbarkeitsformeln beschäftigen, aber auch in Arbeiten aus anderen Bereichen spielen Oberflächenmerkmale eine Rolle. So weisen Biber (1995), Biber/Conrad (2009), Milone (2008), DuBay (2004), Gibson/Levin (1989), Feilke (1996) und Nussbaumer/Sieber (1994) auf die Relevanz von Wort- und Satzlänge hin. Aus den Arbeiten von Kommerell (2008), Christmann/Groebe (2006), Baumann (2009) und Groebe (1982) geht eine klare Präferenz für kurze Wörter und Sätze in Sachtexten hervor.

In einer crosslinguistischen Registerstudie führt Biber (1995:38ff.) einen Sprachvergleich des Englischen mit den drei Sprachen Koreanisch, Nukulaelae Tuvaluisch sowie Somali durch. Dieser Vergleich basiert auf 76 sprachlichen Parametern, die sich bereits in vergangenen Studien als potenziell relevant für das Englische erwiesen haben (vgl. Biber (1995:94ff.)). Unter diesen Parametern wird auch die Wortlänge aufgeführt.

Biber/Conrad (2009:78ff.) stellen ebenfalls eine Liste mit linguistischen Merkmalen zusammen, die sich in vorangegangenen Registerstudien als aussagekräftig für das Englische erwiesen haben. Auch sie führen Wort- und Satzlänge als relevante Faktoren für Registeranalysen auf.

Milone (2008:1ff.) gibt einen Überblick über die Geschichte der Lesbarkeitsformeln. Dabei stellt er Gemeinsamkeiten und Unterschiede bezüglich der sprachlichen Parameter vor, die die jeweilige Formel bestimmen. Vor diesem Hintergrund erläutert Milone (2008:1 u. 3ff.) die Lesbarkeitsformel *Advantage/TASA Open Standard for Readability* (ATOS). Als Hauptparameter nennt Milone (2008:5) unter anderem Wort- und Satzlänge, gemessen in Buchstaben bzw. Wörtern.

DuBay (2004) beschäftigt sich mit der Frage, was Lesbarkeit von Texten ausmacht und wie man diese durch Lesbarkeitsformeln messen kann. Auch er stellt verschiedene Lesbarkeitsformeln und die verwendeten syntaktischen und lexikalischen Parameter genauer vor (vgl. DuBay (2004:13ff.)). In diesem Zusammenhang zeigt DuBay (2004:43f), dass in fast allen Lesbarkeitsformeln die Parameter Wort- und Satzlänge (gemessen in Buchstaben bzw. Wörtern) eine Rolle spielen. Doch ähnlich wie auch Milone (2008:2) warnt DuBay (2004:38) vor einem Trugschluss, denn

„if you just chop up sentences and use shorter words, the results are not likely to improve comprehension. You have to look at the many other factors that affect reading at the level for which you are writing“.

Kommerell (2008) beschäftigt sich mit der Frage, wie man gute unterhaltende Sachtexte für Kinder zwischen 10 und 12 Jahren schreibt. Für meine Arbeit ist besonders relevant,

dass Kommerell sich auch ausführlich mit der sprachlichen Seite der Texte beschäftigt. Sie wendet sich direkt an die Autoren, empfiehlt bestimmte sprachliche Konstruktionen bzw. warnt vor Konstruktionen, die sie für ungeeignet für junge Leser hält und begründet diese auch. Kommerell formuliert diese sprachlichen Merkmale sehr explizit und klar. Sie räumt jedoch ein, dass sie keinen wissenschaftlichen Anspruch erheben könne, da auch sie angesichts der defizitären Forschungslage *aus dem Bauch heraus* und auf der Basis ihrer Erfahrung als Autorin einer Kinderzeitschrift argumentiert (Quelle: persönliches Gespräch). Sie betont, dass es notwendig sei, für junge Leser möglichst kurze Wörter zu verwenden, da sonst der Lesefluss unnötig gestört werde. Sie geht einen Schritt weiter und schreibt einsilbigen Wörtern eine „kräftige, konkrete“ Bedeutung zu (Kommerell (2008:128)), die dem jungen Leser entgegenkommen dürfte (vgl. Abschnitt 3.2.6).

Bezüglich der Satzlänge appelliert Kommerell (2008:132) an den Autor: „Bauen Sie möglichst kurze Sätze!“. Sie hält die Verwendung von kurzen Sätzen für junge Leser aus mehreren Gründen für sinnvoll: Kurze Sätze kommen einer gewissen Ungeduld entgegen, die Kommerell (ebd.) den jungen Lesern unterstellt.

Zudem argumentiert sie mit einem geringeren kognitivem Arbeitsaufwand: Kurze Sätze lassen den jungen Leser das gerade Gelesene nochmals durchdenken und vollständig aufnehmen. Auch zwingen kurze Sätze den Autor dazu,

„einfach und linear zu denken. Sie bringen ihn dazu, treffende und kraftvolle Wörter zu finden, um verständlich und knapp die wichtigsten Informationen [...] auszudrücken. Sie zwingen ihn, alles Überflüssige auch tatsächlich wegzulassen“ (Kommerell (2008:133)).

Kommerell (ebd.) räumt jedoch auch ein, dass der Grundsatz *Kurze Sätze!* nicht bedingungslos gilt, da auch kurze Sätze schwer verständlich sein können. Meiner Meinung nach ist aus genau diesem Grund Kommerells Folgerung, kurze Sätze würden eine einfache Denkweise des Autors garantieren, nicht haltbar.

Gibson/Levin (1989) beschäftigen sich mit dem Sprach- und Leseerwerb von Kindern. In diesem Zusammenhang sehen sie sich die Beschaffenheit von Texten für Kinder genauer an und stellen fest, dass Lesetexte für Fünftklässler sich von denen für Erstklässler und Erwachsene in verschiedenen Punkten sprachlich unterscheiden, unter anderem in Bezug auf die Wortlänge (vgl. Gibson/Levin (1989:137f)). Zudem räumen sie ein, dass der Aspekt *Satzlänge* für die Lesbarkeit eines Textes relevant ist, betonen jedoch, dass ein längerer Satz oft die einfachere Textversion ist. Daher hängt die Lesbarkeit eines Textes nicht von der absoluten Satzlänge ab, sondern eher von der Länge der Teilsätze (vgl. Gibson/Levin (1989:261)).

Christmann/Groeben (2006:151) greifen die Dimensionen des Hamburger Verständlichkeitskonzepts in leicht veränderter Form auf als

- (a) sprachliche Einfachheit,
- (b) kognitive Gliederung/Ordnung,
- (c) Kürze/Prägnanz und
- (d) motivationale Stimulanz.

Auch wenn für Christmann/Groeben (2006: 151f) die Merkmale der Dimension (b) „am gewichtigsten für die Verständlichkeit“ sind, unterstützen die Merkmale der Dimension (a)

„eine unaufwendige und reibungslose Textrezeption“ (Christmann/Groeben (2006:154)), welche grundlegend für Verständlichkeit ist.

Hier betonen Christmann/Groeben (ebd.), dass

„sich auf Wortebene insbesondere der Gebrauch kurzer [...] Wörter [...] als verarbeitungserleichternd erwiesen [hat] [...] Bezüglich der syntaktischen Einfachheit [...] gilt als gesichert, dass [...] überlange Sätze die Verarbeitung erschweren“.

Christmann/Groeben (2006:152 u. 154) berufen sich mit ihren Annahmen auf die empirische Forschung: Alle Annahmen wurden experimentell nachgewiesen. Dies ist somit die einzige Arbeit aus dem Bereich der Sachtextforschung, die mir wissenschaftlich nachgewiesene sprachliche Merkmale liefern kann.

Auch Baurmann (2009:34f) betont im Zusammenhang mit den Dimensionen des Hamburger Verständlichkeitskonzepts, dass die Wortlänge die Verständlichkeit eines Textes beeinflusst: Ein Autor sollte demnach eher kurze Wörter verwenden. Entsprechend wirkt es sich auch positiv auf die Verständlichkeit aus, wenn der Autor „kurze und grammatisch einfache Sätze verwendet“ (ebd.).

Groeben (1982) beschäftigt sich mit der Verständlichkeit von literarischen und Informationstexten. Er versucht, Verständnisbeeinflussende Merkmale eines Textes „möglichst objektiv, quasi von außen, festzustellen, indem man sich auf bestimmte Aspekte der sprachlichen Oberflächenstruktur konzentriert“ (vgl. Groeben (1982:173)). Diese "material-objektive[n] Textmerkmale" (ebd.) spielen in Lesbarkeitsformeln eine große Rolle. Er stellt die "Relevanz eines Wort- und eines Satz factors" (Groeben (1982:185)) als Gemeinsamkeit aller Lesbarkeitsformeln heraus und leitet daraus sprachliche "Regeln für das Schreiben von Informationstexten mit einer möglichst lesbaren sprachlich-stilistischen Form" ab (ebd.). Für die Dimension *sprachliche Einfachheit* im Rahmen des Hamburger Verständlichkeitskonzepts formuliert Groeben (ebd.) die Anforderung, „kurze Worte (nach Buchstaben und/oder Silbenlänge)“ zu verwenden. Zudem plädiert er für eine reduzierte Satzkomplexität durch die „Verwendung möglichst kurzer und grammatikalisch einfacher Sätze“ (ebd.).

Feilke (1996:200) untersucht die Entwicklung syntaktischer Fähigkeiten in geschriebenen Texten von Schülern aus den Klassen 4, 7, 10 und 12 im Vergleich zu Studenten. Seine Probanden sind zwischen 10 und 22 Jahren alt. Er vergleicht die Texte in seiner Aufsatzstudie unter anderem in Bezug auf die Satzlänge (gemessen in Wörtern). Somit geht auch er davon aus, dass die Satzlänge ein relevanter sprachlicher Parameter beim Textvergleich ist.

Nussbaumer/Sieber (1994:149) stellen das *Zürcher Textanalyseraster* vor, mit dem es möglich ist, einen Text auf verschiedenen Ebenen zu analysieren. Das Raster ist

„eine systematische Zusammenstellung von einzelnen Fragen an einen Text [...], von Fragen, in die sich die ganz generelle Frage ‚Wie ist der Text?‘ auseinandernehmen lässt“.

Allgemeine Merkmale und Auffälligkeiten von Texten sollen durch die verschiedenen Kategorien des Rasters erfasst werden, damit die Texte objektiv charakterisiert und schließlich wissenschaftlich analysiert werden können (Nussbaumer/Sieber

(1994:149f)). Auch in diesem Modell spielt die Satzlänge eine Rolle, indem sie als allgemeines Merkmal der Texte erfasst wird und so in die Charakterisierung der Texte einbezogen wird (vgl. Nussbaumer/Sieber (1994:153)). Die Satzlänge ist eines der Merkmale, die im Raster als „Bezugsgrößen, als Korrelate für festgestellte Stark- und Schwachstellen des Textes unbedingt mit in Betracht gezogen werden müssen“ (vgl. Nussbaumer/Sieber (1994:156)).

Aus den hier betrachteten Arbeiten geht hervor, dass folgende Merkmale in meiner Arbeit relevant sein können:

- (1) Wortlänge in Buchstaben
- (2) Satzlänge in Wörtern

3.2.2 Wortbildung

Im Zusammenhang mit der Wortlänge ist es sinnvoll, das Merkmal *Wortbildung*, genauer *Komposition*, zu betrachten. Das Deutsche ist diesbezüglich eine sehr produktive Sprache (vgl. auch Eisenberg (2004:226)). Gierlich (2005:32) bemerkt, dass Komposita in Sachtexten gehäuft auftreten. Sie seien Ausdruck einer Tendenz zur sprachlichen Ökonomie. Diese wiederum führt zu einer sehr kompakten und komprimierten Sprache, die schwer verständlich ist.

In meiner Arbeit kann demnach das folgende Merkmal relevant sein:

- (3) Anzahl der verwendeten Komposita

3.2.3 Adjektivfrequenz

May (1996) und Kommerell (2008) setzen sich mit der Frage auseinander, in welchem Umfang man Adjektive in Sachtexten für junge Leser einsetzen sollte.

Nach May (1996:44) zeichnet sich ein sprachlich angemessener Text unter anderem dadurch aus, dass der Leser ihn als „reichhaltig und bildhaft“ wahrnimmt. Dies erreicht ein Autor unter anderem durch eine „reiche[n] Verwendung von Adjektiven“ (ebd.).

Kommerell (2008:129) sieht im Gegensatz zu May in der Verwendung zahlreicher Adjektive eine deutliche Qualitätsminderung eines Textes, da dieser schnell wie ein Schulaufsatz wirke. Ein sparsamer Einsatz von Adjektiven hingegen würde für einen guten Text sprechen, denn „die meisten Eigenschaftswörter vermitteln unter Kommunikationsgesichtspunkten keine Informationen“ (ebd.).

Aus den hier betrachteten Arbeiten geht hervor, dass das folgende Merkmal in meiner Arbeit relevant sein kann:

- (4) Anzahl der verwendeten Adjektive

3.2.4 Verhältnis von attributiven zu prädikativen Adjektiven

Biber (1995) und Doderer (1960) gehen genauer auf den Adjektivgebrauch ein und betrachten den Gebrauch von attributiven Adjektiven im Unterschied zu prädikativen Adjektiven.

Aus der Liste der 76 sprachlichen Parameter, die sich nach Biber (1995:94) bereits in vergangenen Studien für das Englische als potenziell relevant erwiesen haben, geht hervor, dass die Verwendungshäufigkeit von attributiven im Vergleich zu prädikativen Adjektiven eine Register unterscheidende Rolle spielen kann (vgl. Biber (1995:96), Biber (2009:830), Biber/Conrad (2009:80)). Dies ist auch relevant im Hinblick auf die Frage, ob in *GEOLino* ein bestimmtes sprachliches Register vorliegt.

Auch bei Doderer (1960) finden sich Hinweise darauf, dass die Verwendung von attributiven Adjektiven relevant für meine Arbeit sein kann. Doderer stellt sich die Frage, wie Kinder erzählen und unter Berücksichtigung welcher Gesetzmäßigkeiten sie das tun. In einer empirischen Studie vergleicht er die Erzählungen von Schulanfängern mit denen von 8-10jährigen Schülern. Dabei stellt er eine Entwicklung der Erzählfähigkeiten im Rahmen des Spracherwerbs fest. Er beobachtet bei den 8-10jährigen Schülern unter anderem, dass „Eigenschaftswörter [...] häufiger als vorher attributiv gebraucht [werden]“ (Doderer (1960:321)).

Aus den hier betrachteten Arbeiten geht hervor, dass das folgende Merkmal in meiner Arbeit relevant sein kann:

- (5) Anzahl der verwendeten attributiven Adjektive bzw. Verhältnis von attributiven zu prädikativen Adjektiven

3.2.5 Komparation

Kommerell (2008) und Gierlich (2005) stellen die Frage nach dem Gebrauch von Komparation in Sachtexten und kommen zu entgegengesetzten Ergebnissen.

Im Zusammenhang mit dem Adjektivgebrauch betont Kommerell (2008:129), dass in Texten für junge Leser Superlative vermieden werden sollten, da diese „meist ohne Substanz“ seien. Sie hält den Einsatz von Vergleichen an dieser Stelle für aussagekräftiger.

Gierlich (2005:28f) hingegen sieht (in Anlehnung an Niederhauser (1999)) den Einsatz von Komparation als eine Strategie der „popularisierende[n] Wissensvermittlung“, die einen Text für fachexternes Publikum zugänglich und verständlich macht.

Aus den hier betrachteten Arbeiten geht hervor, dass das folgende Merkmal in meiner Arbeit relevant sein kann:

- (6) Komparation und speziell Gebrauch des Superlativs gegenüber Gebrauch des Positivs und Komparativs

3.2.6 Abstrakta

In einigen Arbeiten findet man Hinweise auf den Gebrauch von abstrakten Begriffen. So weist DuBay (2004) allgemein auf die Relevanz dieses Merkmals hin. Kommerell (2008), Baurmann (2009), Christmann/Groeben (2006) und Groeben (1982) gehen in ihren Arbeiten näher auf die Frage ein, wie sich der Gebrauch von Abstrakta in Sachtexten auf den Leser auswirkt und kommen zu übereinstimmenden Ergebnissen.

In Anlehnung an Klare führt DuBay (2004:39) eine Liste mit Merkmalen an, die das Verstehen eines Textes beeinflussen. Hier nennt er auch das Verhältnis von konkreten zu abstrakten Wörtern.

Kommerell (2008:130) bevorzugt die Verwendung von konkreten Substantiven für junge Leser: „Die meisten bildnahen Substantive erwecken ebenfalls noch konkrete Vorstellungen: Schlaf, Neid, Ferien, Klang“ (ebd.). Vom Gebrauch abstrakter Substantive rät sie jedoch ab, denn „bidleere und geblähte Hauptwörter haben in einem Text für junge Leser nichts zu suchen“ (ebd.). Substantive mit den Suffixen *-ung*, *-ismus*, *-heit* und *-keit* fallen nach Kommerell (ebd.) in diese Kategorie der „erklärungsbedürftige[n] Begriffe“.

Auch Christmann/Groeben (2006:154) betonen, dass sich nicht nur kurze, sondern auch konkrete Wörter verarbeitungserleichternd auswirken.

Ähnlich wie Christmann/Groeben nimmt Baurmann (2009:34) an, dass neben der Wortlänge auch der Abstraktionsgrad der Wörter das Verstehen beeinflusst. Er fordert, dass eher „konkrete und anschauliche Wörter“ (ebd.) verwendet werden sollten, um die Verständlichkeit eines Textes zu optimieren.

Groeben (1982:225) betont die „lern- und behaltensfördernde Wirkung von konkreten, anschaulichen Worten“. Diese Wirkung beruht darauf, dass der Leser durch konkrete Wörter schneller auf das entsprechende Konzept im mentalen Lexikon zugreifen kann (ebd.). Um den Abstraktionsgrad von Wörtern – hier Substantive – festlegen zu können, verweist Groeben (1982:226), wie Kommerell, auf bestimmte Suffixe: *-heit*, *-ie* ([i:]), *-ik*, *-ion*, *-ismus*, *-ität*, *-keit*, *-nz*, *-tur* sowie *-ung* kennzeichnen abstrakte Substantive. In diesem Zusammenhang stellt Groeben (ebd.) ein Verfahren vor, um den „Abstraktheitsindex“ des Textes zu berechnen. Dieses Verfahren greife ich in Abschnitt 5.3.2 im Rahmen der nominalen Abstrakta auf.

Aus den hier betrachteten Arbeiten zum Gebrauch von abstrakten Begriffen geht hervor, dass das folgende Merkmal in meiner Arbeit relevant sein kann:

(7) Verhältnis von abstrakten zu konkreten Substantiven

3.2.7 Konjunktionsfrequenz

Willenberg (2005) und Groeben (1982) untersuchen den allgemeinen Gebrauch von Konjunktionen in Sachtexten und kommen zu entgegengesetzten Ergebnissen, was die Verständnisfördernde Wirkung von Konjunktionen betrifft.

Willenberg (2005) beschäftigt sich mit dem Problem, den Schwierigkeitsgrad eines Textes einzuschätzen. Er betont, dass man einen Fehler macht, wenn man einen Text allein aufgrund seines Inhalts einschätzt statt sich die formale Seite des Textes anzusehen (vgl. Willenberg (2005:94)). Er spricht sich gegen vergangene Versuche aus, Texte formal zu kategorisieren. Lesbarkeitsformeln zum Beispiel, die größtenteils im Rahmen des Englischen entwickelt wurden, lassen sich seiner Ansicht nach nicht einfach auf andere Sprachen übertragen (ebd.). Daher entwickelt Willenberg für das Deutsche eigene Kriterien, die für die Beurteilung eines Textes relevant sein sollten. Unter diesen Kriterien findet sich auch der Einsatz von „Junktoren“ (Willenberg (2005:95)), die unter anderem als Konjunktionen auftreten können. Diese fördern nach Willenberg (2005:97f) die logische Kohärenz eines Textes und erleichtern das Verständnis.

Groeben (1982:218) weist im Zusammenhang mit dem Aspekt *Textverständlichkeit* auf „Techniken der Textoptimierung“ hin. Diese Techniken sollen dazu beitragen, einen Text besser verständlich zu machen. Auch bei Groeben spielen Konjunktionen eine Rolle: Durch Studien konnte experimentell nachgewiesen werden, dass kürzere Sätze signifikant besser

verstanden und behalten wurden, auch wenn der absolute Unterschied nicht groß war. Aus diesen Erkenntnissen schließt Groeben, dass (satzverknüpfende) Konjunktionen im Zuge der Textoptimierung vermieden werden sollten, da sie die Sätze per se in die Länge ziehen. Doch auch hier vermutet Groeben (1892:230), dass der absolute Unterschied nicht groß ist. Potenziell schätzt er die Verständnisfördernde Wirkung von Konjunktionen als negativ ein.

Aus den hier betrachteten Arbeiten geht hervor, dass das folgende Merkmal in meiner Arbeit relevant sein kann:

- (8) Allgemeine Frequenz von Konjunktionen

3.2.8 Semantische Klassen von Konjunktionen

Die Arbeiten von Hiller (2010) und Feilke (1996) legen nahe, dass der Einsatz von adversativen und kausalen Konjunktionen eine besondere Bedeutung für junge Leser hat.

Hiller (2010) beschäftigt sich mit der Frage, welche Einzelprozesse beim Lesen und Verstehen von Sachtexten ablaufen und welche leser- und textbezogenen Faktoren den Leseprozess und -erfolg beeinflussen können. Als einen wichtigen Faktor hebt er dabei auch die sprachliche Gestaltung des Textes hervor, die das Verstehen beeinflusst (vgl. Hiller (2010:43f)). Hiller sieht vor allem in kausalen, aber auch in adversativen Verknüpfungen (zum Beispiel in Form von entsprechenden Konjunktionen) Verständnisfördernde sprachliche Gestaltungsmittel, die die Kohärenz eines Sachtextes schaffen (vgl. Hiller (2010:44)). Verwendet ein Autor demnach kausale und adversative Verknüpfungen, wird der Leser den Text schneller rezipieren und besser behalten können. Was Willenberg für Konjunktionen im weiteren Sinne feststellt, stellt Hiller in Bezug auf kausale und adversative Verknüpfungen fest.

Auch Feilke (1996) greift kausale Konjunktionen als sprachlichen Parameter in seiner vergleichenden Aufsatzstudie auf. Dabei bemerkt er, dass die Schüler ab einem Alter von ca. 12 Jahren Konjunktionen seltener verwenden als die jüngeren Schüler; besonders ausgeprägt sei diese Entwicklung bei kausalen Konjunktionen, deren Verwendung bereits mit 10 Jahren abnimmt (vgl. Feilke (1996:203)).

Aus den hier betrachteten Arbeiten geht hervor, dass folgende Merkmale in meiner Arbeit relevant sein können:

- (9) Gebrauch von adversativen Konjunktionen
- (10) Gebrauch von kausalen Konjunktionen

3.2.9 Leserkontakt

In fast allen von mir betrachteten Arbeiten zu Sachtexten für junge Leser spielt der Leserkontakt eine wesentliche Rolle. In ihren Arbeiten schreiben Baurmann (2009), Göpferich (2008), Kommerell (2008), Schulz/Tausch/Langer (2011) und May (1996) dem Leserkontakt einen positiven Effekt zu. Hier lässt sich also ein genereller Konsens beobachten. Biber/Conrad/Reppen (2002), Biber (2006) sowie auch Biber/Conrad (2009) gehen außerdem der Frage nach, ob der direkte Leserkontakt eine Register unterscheidende Funktion hat.

Baurmann (2009) spricht sich für eine Textbetrachtung aus, die auch den sprachlichen Aspekt einschließt. In diesem Zusammenhang analysiert er einen Sachtext aus einem Schulbuch und stellt fest:

"Sprachlich kommt der Text den Kindern entgegen, ist er doch in Wortwahl und Satzbau ihrem Sprachgebrauch nahe [...] Darüber hinaus wird es jugendliche [...] Leser ansprechen, dass sie mehrfach direkt angesprochen werden." (Baurmann (2009:19))

Baurmann sieht in der direkten Anrede an den Leser ein Merkmal, das einen Sachtext aufwertet und interessant macht.

Göpferich (2008) stellt die Frage, was einen verständlichen Text ausmacht und wie man Textverständlichkeit messen kann. Nach dem Hamburger Verständlichkeitskonzept wird die Verständlichkeit eines Textes von den vier Dimensionen

- (a) Einfachheit (in Bezug auf die sprachlichen Gestaltungsmittel)
- (b) Gliederung/Ordnung
- (c) Kürze/Prägnanz und
- (d) Anregende Zusätze

beeinflusst (vgl. Schulz/Langer/Tausch (2011:19)).

Göpferich(2008:295 u. 302) stellt das Karlsruher Verständlichkeitskonzept als Erweiterung des Hamburger Verständlichkeitskonzepts vor. Beide Konzepte gehen davon aus, dass es möglich ist, Texte auf objektiver sprachlicher Ebene zu beschreiben und zu beurteilen. In beiden Konzepten gibt es verschiedene Dimensionen, die die Verständlichkeit eines Textes beeinflussen. Eine dieser Dimensionen betrifft „anregende Zusätze bzw. motivationale Stimulanz“ (Göpferich (2008:295)) im Hamburger Konzept bzw. „Motivation“ (Göpferich (2008:302)) im Karlsruher Konzept. Göpferich (2008:304) versteht Motivation hier textintern als Motivation, „die der Text aus sich heraus schafft“; im Gegensatz zur textexternen Motivation, die vom Leser ausgeht. In diesem Zusammenhang hat der Text unter anderem die Aufgabe, das Interesse des Lesers zu wecken und aufrecht zu erhalten sowie die Lesemotivation zu erhöhen. Dies gelingt dem Autor zum Beispiel dadurch, dass er „den Leser persönlich anspricht“ (Göpferich (2008:305)).

Auch Kommerell (2008:145) sieht in der direkten Anrede an den jungen Leser ein geeignetes Stilmittel, denn „Die direkte Ansprache mit ‚du‘ und ‚ihr‘ bietet jungen Lesern den Zugang zur persönlichen Betroffenheit, der Texte so attraktiv macht“.

Schulz/Tausch/Langer (2011:21ff.) stellen fest, dass sich Texte anhand der Merkmale des Hamburger Verständlichkeitskonzepts gut beurteilen lassen. In diesem Zusammenhang betonen sie (2011:161 u. 165), dass die Achtung bzw. Rücksichtnahme des Autors auf den Leser ein notwendiges Merkmal verständlicher Texte ist. Diese Rücksichtnahme drückt sich unter anderem dadurch aus, dass der Autor in seinem Text eine Beziehung zum Leser aufbaut, indem er einen Kontakt herstellt (vgl. Schulz/Tausch/Langer (2011:165)). Auch hier spielt also die direkte Kontaktaufnahme des Autors zum Leser eine wesentliche Rolle.

Auch nach May (1996:42) ist der „Leserbezug“ eine wichtige Kategorie, nach der Texte beurteilt werden können. Dieser kann unter anderem durch die „direkte Ansprache des

Lesers“ (ebd.) hergestellt werden. Nach May (1996) ist der Leserkontakt ein relevanter Faktor, der einen Text positiv beeinflusst.

Biber/Conrad/Reppen (2002) führen eine Registerstudie durch und untersuchen den Gebrauch von verschiedenen Referenz-Typen in jeweils zwei gesprochenen und geschriebenen Registern des Englischen: „conversation and public speeches [...] and news reportage and academic prose“ (Biber/Conrad/Reppen (2002:110)).

In diesem Rahmen ist der Referenz-Typ interessant, den Biber/Conrad/Reppen (2002:111) „exophoric referents“ (ER) nennen. Darunter sind Pronomen zu verstehen, die sich auf textexterne Elemente beziehen¹²: „the pronouns *I* and *you* refer directly to the speaker and addressee“ (ebd.).

Ein Vergleich der vier Register zeigt einen signifikanten Unterschied in der Verwendung der ER zwischen den gesprochenen und den geschriebenen Registern, denn „exophoric pronouns account for over half of all given references in conversation“ (Biber/Conrad/Reppen (2002:117)). Auch im zweiten gesprochenen Register – *public speeches* – sind ER stark vertreten: Sie machen ungefähr ein Drittel der verwendeten referenz-Typen aus (vgl. Biber/Conrad/Reppen (2002:118)).

Im Gegensatz dazu zeigen die Ergebnisse, dass

„this reliance on exophoric reference is not found in either written register; in fact, apart from an occasional reference to the author (*I*), exophoric references are almost completely absent from these written registers. Instead, the written registers have a greater reliance on anaphoric references“ (ebd.).

Die Studie ergibt also, dass ER durchaus typisch für gesprochene Register sind, für geschriebene Register jedoch sehr untypisch. Somit wirkt dieses Merkmal hier Register unterscheidend.

Eine ähnliche Beobachtung macht Biber (2006) in einer vergleichenden Studie zu geschriebenen und gesprochenen Sub-Registern innerhalb des Registers *Universitätssprache*. Er vergleicht jeweils fünf gesprochene und geschriebene Sub-Register unter anderem in Bezug auf die Verwendungshäufigkeit von Personalpronomen (vgl. Biber (2006:51)). Dabei stellt er fest, dass „spoken university registers rely on pronouns to a greater extent than the written registers“ (ebd.).

Auch in der Liste von Biber/Conrad (2009:79) finden sich entsprechende Hinweise auf eine Register unterscheidende Funktion von Personalpronomen.

Aus den hier betrachteten Arbeiten geht hervor, dass das folgende Merkmal in meiner Arbeit relevant sein kann:

(11) Leserkontakt

Nachdem ich die Merkmale herausgearbeitet habe, möchte ich noch kurz darauf eingehen, auf welchem kognitiven Entwicklungsstand sich die *GEOLINO*-Leser im Allgemeinen befinden.

¹² Im Gegensatz dazu beziehen sich anaphorische Pronomen auf textinterne Elemente (vgl. Biber/Conrad/Reppen (2002:111)).

3.3 Psycholinguistische Grundlage: Kognitiver Entwicklungsstand

Da die Zielgruppe der *GEOLino* keine Leseanfänger mehr sind, möchte ich kurz erläutern, in welchem Stadium des Leseerwerbs sich die jungen Leser im Durchschnitt befinden. Dabei gehe ich grundsätzlich von Kindern aus, die eine ihrem Alter entsprechende Entwicklung zeigen und zwar auf allen kognitiv-psychologischen Ebenen. Über die einzelnen Prozesse und Entwicklungsstufen im Leseerwerb gibt es zahlreiche Erklärungsansätze. Eine erschöpfende Diskussion und Gegenüberstellung kann ich im Rahmen dieser Arbeit nicht leisten. Ich richte mich im Wesentlichen nach dem Ansatz von Bredel/Fuhrhop/Noack (2011) und konzentriere mich auf die zentralen Leseerwerbsphasen in alphabetischen Schriftsystemen.

Nach Bredel/Fuhrhop/Noack (2011:91) ist der erste und wichtigste Schritt im Leseerwerb die Entwicklung der phonologischen Bewusstheit (vgl. auch Ehri (2005:173), Veirhoeven/Reitsma/Siegel (2010:389)). Hier lernen Kinder, dass der Sprachfluss der gesprochenen Sprache in einzelne Wörter zerlegt werden kann, diese wiederum in einzelne Silben und Phoneme. Phonologische Bewusstheit ist die Voraussetzung dafür, dass Kinder als Leseanfänger Buchstaben bestimmte Laute zuordnen können. So erlernen sie auch die systematische Zuordnung zwischen Graphemen und Lauten. Auf dieser Ebene des Leseerwerbs stehen „die basalen Teilprozesse, also die Entschlüsselung der schriftsprachlichen Zeichen“ (Bredel/Fuhrhop/Noack (2011:121)) im Vordergrund.

Erst wenn die Kinder auf dieser Ebene sicher genug sind, kann die zweite große Phase – das automatisierte Lesen – einsetzen, welches nach Bredel/Fuhrhop/Noack (2011:121) „das sinnkonstruierende, verstehende und interpretierende Lesen“ oder nach Rosebrock/Nix (2008:74) „reading to learn“ ist. Der Erwerb des automatisierten Lesens braucht jedoch mehrere Jahre (Bredel/Fuhrhop/Noack (2011:129)). Wann das automatisierte Lesen ungefähr einsetzen sollte, kann man aus dem Rahmenplan für den Deutschunterricht schließen: Man geht hier davon aus, dass die Kinder am Ende der vierten Klasse in der Lage sein müssen, (kritisch) mit Texten umzugehen:

„[Sie] unterscheiden literarische Texte und Sachtexte, [...] entnehmen Texten gezielt Informationen, ziehen Schlussfolgerungen aus Texten, äußern ihre Gedanken und Meinungen zu Texten“¹³.

Dies stimmt mit Rosebrock/Nix (2008:74) Annahme überein, dass ab der 5. Klasse das Lernen und nicht mehr das Lesen lernen im Vordergrund steht.

Rayner/Pollatsek (1989) zeichnen die einzelnen Schritte des Leseerwerbs vom Kleinkindalter an nach. Auch sie sehen im Übergang von der vierten zur fünften Klasse einen wichtigen Entwicklungsschritt, denn ab der fünften Klasse lesen Kinder auf Erwachsenen-Niveau (1989:391).

Prinzipiell kann man also sagen, dass *GEOLino*-Leser als ältere Grundschulkinder die grundlegenden Entwicklungsschritte im Leseerwerb bereits hinter sich gebracht haben. Was sie von Lesern auf Erwachsenen-Niveau unterscheidet, mag einerseits eine noch nicht ganz robuste Routine im Lesen sein, da die langjährige Übung und Leseerfahrung fehlt. Andererseits betonen Bos et al. (2003:78f), dass Lesekompetenz durch zwei Hauptkomponenten bestimmt wird: die Fähigkeit, textimmanente Information zu nutzen und die Fähigkeit, externes Wissen einzusetzen, um den Text zu verstehen und zu

¹³<http://www.bildungsserver-mv.de/download/rahmenplaene/rp-deutsch-gs.pdf> vom 9.2.12.

bewerten. Es liegt auf der Hand, dass Grundschüler nicht über das gleiche Maß an externem Wissen/Weltwissen verfügen wie Erwachsene. Daher können sie dieses Wissen nur in geringerem Maß zum Textverstehen heranziehen.

Bezüglich des Spracherwerbs-Standes kann man sagen, dass ältere Grundschulkinder auch hier bereits die wesentlichen Stufen durchlaufen haben. Dietrich (2007:72) spricht vom „fließenden Verstehen und Produzieren von längeren Gesprächsbeiträgen im Alter zwischen 10 und 12 Jahren“. Man kann keine absolute Grenze nennen, also ein Alter angeben, ab dem ein Mensch mit dem Spracherwerb „fertig“ ist, denn Spracherwerb dauert – jedenfalls auf der lexikalischen Ebene – das ganze Leben an. Die von mir gelesenen Arbeiten zum Spracherwerb gehen jedoch implizit davon aus, dass der Erwerb von lexikalischen und syntaktischen Fähigkeiten schon in einer sehr frühen Phase grundsätzlich gefestigt ist. In diesen Arbeiten werden nämlich Kinder betrachtet, die altersmäßig weit vor dem *GEolino*-Alter liegen (vgl. Dijkstra/Kempen (1993), Tracy (2008), Lust (2009), Dietrich (2007)). So ist zum Beispiel das älteste Kind, von dem Tracy (2008:185) berichtet, fünf Jahre alt. Schon Deußing (1927:86f) beobachtet, dass die lexikalische Entwicklung von Kindern mit einer sehr ungleichmäßigen Altersverteilung dokumentiert wird: Die meisten Aufzeichnungen betreffen den aktiven Wortschatz von 1,5- bis 2,5jährigen Kindern. Deutlich weniger Aufzeichnungen gibt es für 6jährige und ältere Kinder. Dies spricht auch dafür, dass Kinder bereits im Vorschulalter bzw. beim Eintritt in die Grundschule die wesentlichen Erwerbsstufen durchlaufen haben. Daher gehe auch ich davon aus, dass die Zielgruppe von *GEolino* mit 10 bis 12 Jahren die wichtigsten Erwerbsschritte hinter sich gebracht hat.

Was bezüglich des Spracherwerbs noch nicht ausgereift ist, liegt am ehesten im Bereich der pragmatischen Kompetenz und betrifft zum Beispiel das Verstehen von Ironie und Metaphern (vgl. Meibauer (2008:169), Johnson/Leone (1989:1f), Dijkstra/Kempen (1993)). Nach Meibauer (2008:169) ist die „spätere pragmatische Entwicklung [...] auch von dem immer größer werdenden Situations- und Weltwissen abhängig“. Hier liegt ein ganz ähnlicher Sachverhalt vor, wie ich ihn im Zusammenhang mit dem Leseerwerbs-Status beschrieben habe: Der Unterschied zwischen Kindern und Erwachsenen besteht darin, dass den Kindern ein ausgeprägtes externes Wissen/Weltwissen fehlt. Offenbar werden pragmatische Fähigkeiten erst über viele Jahre hinweg vollständig erworben; nach Meibauer (2008:174) erstreckt sich der Erwerb der pragmatischen Kompetenz „praktisch über die gesamte Schulzeit“.

Dies soll lediglich ein Überblick über den kognitiven Entwicklungsstand der *GEolino*-Leser sein. Festhalten möchte ich hier, dass die *GEolino*-Leser den *GEO*-Lesern durch noch nicht ausgereiftes Weltwissen entwicklungsbedingt nachstehen. Dies wirkt sich auch auf das Verstehen von gelesenen Texten aus. Die grundsätzliche Kompetenz, Texte zu lesen, ist jedoch bei den *GEolino*-Lesern bereits vorhanden.

3.4 Exkurs 2: Registerbegriff

In meiner Fragestellung (2) habe ich den Begriff *Register* angesprochen. Ich möchte hier kurz klären, was darunter zu verstehen ist. Die Grundidee des Registerbegriffs – und das ist auch der relevante Aspekt in meiner Arbeit – ist die sprachliche Varianz. Ein Sprecher hat immer mehrere Möglichkeiten, einen Gedanken zu äußern und trifft unbewusst eine Wahl, welche sprachlichen Mittel er dafür verwendet. Es liegt an ihm, ob er zum Beispiel eine Aktiv- oder eine Passivkonstruktion verwendet, einen Satz als Nebensatz einbettet statt ihn

als Hauptsatz anzureihen oder eine erklärende Umschreibung statt des Fremdwortes benutzt (vgl. auch Biber (2009:823)). Biber (ebd.) definiert *Register* als

"a cover term for any language variety defined by its situational characteristics, including the speaker's purpose, the relationship between speaker and hearer, and the production circumstances".

Ein Register ist also die Sprachverwendung in einer bestimmten Situation (vgl. Biber/Conrad (2009:6)). Jedes Register weist dabei bestimmte Strukturen und Merkmale mit einer spezifischen Frequenz auf, die es als solches charakterisiert und von anderen Registern unterscheidet (vgl. Ferguson (1994:20), Biber/Conrad (2009:6), Biber (2009:823)).

Ein Register lässt sich anhand von Variablen oder *linguistic features* beschreiben (vgl. Biber (1995:29)). Die Variable *Satzmodus* zum Beispiel kann einen der beiden Werte *Aktiv* und *Passiv* annehmen, je nachdem, in welchem sprachlichen Register man sich bewegt. Hier lassen sich Register direkt miteinander vergleichen und voneinander unterscheiden.

Bezogen auf meine Fragestellung (2) bzw. Hypothese (1) würde das bedeuten, dass die Variablen im Register *GEolino* andere Werte (= Varianten) aufweisen als im Register *GEO*.

4. Methode

Anhand der in Abschnitt 3.2 ermittelten Merkmale möchte ich mit einer Korpusstudie eventuell vorhandene Unterschiede zwischen *GEolino* und *GEO* herausarbeiten. Dazu führe ich einen Vergleich durch, in welchem ich ein Textkorpus mit *GEolino*-Texten sowie ein Vergleichskorpus mit *GEO*-Texten erstelle und beide Korpora gegenüberstelle. Einzelheiten zur Korpuserstellung erläutere ich im folgenden Abschnitt 4.1. Ich annotiere die Korpora gemäß den ermittelten und ausgewählten sprachlichen Merkmalen. Dazu formuliere ich in Abschnitt 4.2 entsprechende Richtlinien.

Ich vergleiche die beiden Korpora auf quantitativer Basis, das heißt ich ermittle, mit welcher Frequenzverteilung die sprachlichen Merkmale in beiden Korpora auftreten und wie sich diese Verteilung unterscheidet. Dadurch sollen folgende Fragen beantwortet werden:

- (1) Treten die Merkmale, die in der Forschungsliteratur als unangemessen oder untypisch für Sachtexte für junge Leser eingestuft werden, in der *GEolino* nur marginal, in der *GEO* hingegen häufiger auf?
- (2) Treten die Merkmale, die in der Forschungsliteratur als angemessen oder typisch für Sachtexte für junge Leser eingestuft werden, in der *GEolino* häufig, in der *GEO* hingegen marginal auf?

Eine statistische Analyse soll zeigen, ob sich die Korpora bezüglich der untersuchten Merkmale signifikant voneinander unterscheiden.

Dazu führe ich zunächst eine Hauptkomponentenanalyse durch. Hier kann man sehen, ob und inwiefern sich eine Clusterung (eine sprachliche Gruppierung) der Subkorpora zeigt. Wenn eine Clusterung zu erkennen ist, würde dies bedeuten, dass sich die Texte auf der

sprachlichen Ebene systematisch voneinander unterscheiden. Dies würde wiederum ein Indiz für unterschiedliche Register sein.

In einem univariaten Vergleich ermittle ich schließlich die Auftretenshäufigkeit der sprachlichen Merkmale in den Korpora und stelle gegebenenfalls auch einzelne Merkmale gegenüber.

4.1 Korpora: Textauswahl und Variablen

Es gibt eine Vielzahl von Punkten, die man bei der Erstellung von Korpora beachten muss. Aus Platzgründen kann ich hier nicht erschöpfend darauf eingehen, welche generellen Vorüberlegungen zum Korpusdesign anzustellen sind. Zu Aspekten wie *Repräsentativität*, *Größe* und *Ausgewogenheit* eines Korpus verweise ich daher auf Hunston (2008), Scherer (2006) und Lemnitzer/Zinsmeister (2010).

Ich lege für die Korpuserstellung folgende Kriterien fest, nach denen ich die Texte auswähle:

- (1) Gesamt-Tokenzahl von ca. 10.000 Token pro Korpus (dabei ungefähr gleich lange Texte)
- (2) Themengleichheit der Texte
- (3) möglichst gleiche Verteilung der Metadaten zu den Autoren (Gleichverteilung von *männlich* und *weiblich*)
- (4) pro Subkorpus nur ein Text je Autor

Die Kriterien (2) bis (4) erläutere ich in den folgenden Abschnitten genauer. Zunächst ist es jedoch wichtig, auf die Begriffe *Unabhängige Variable* (UV), *Abhängige Variable* (AV) und *Störvariable* (SV) einzugehen.

Die Begriffe erklären sich durch ihre Relation zueinander: Nach Meindl (2011:34) hat eine UV einen Einfluss auf eine AV. Sie kann vom Versuchsleiter manipuliert werden, wodurch sich ihr Einfluss auf die AV verändern kann. Die UV ist also „die vermutete Einflussgröße oder Ursache, die ‚Schraube, an der gedreht wird‘“ (ebd.). Die AV hingegen spiegelt die Manipulation an der UV wider und verändert sich relativ dazu. Die AV „soll die Wirkung der UV erfassen [...] Sie hängt von der UV ab“ (ebd.).

Bezogen auf meine Arbeit heißt die UV *Korpus*. Die Manipulation ist hier im weiten Sinne zu verstehen und bedeutet einfach die Ausprägungen als *GEOLino* und *GEO*. Laut Hypothese (1) – Die Sprache in der *GEOLino* unterscheidet sich systematisch von der Sprache in der *GEO* – beeinflusst die Ausprägung der UV *Korpus* die Wahl der sprachlichen Mittel. Die AV ist hier also die quantitative Ausprägung (Frequenz) der einzelnen sprachlichen Merkmale.

Ich kann jedoch nicht ausschließen, dass es weitere Variablen gibt, die neben der UV einen Einfluss auf die AV haben. Sprachliche Unterschiede müssen nicht zwangsläufig durch die UV *Korpus* bedingt sein, die ich hier festlege. „Alle Variablen, die einen Einfluss auf die AV haben (mit Ausnahme der durch den Versuchsleiter willkürlich manipulierten UV)“ (Meindl (2011:35)), sind Störvariablen (SV).

Nach Meindl (2011:39) kann man den Einfluss von SV auf die AV auf mehrere Arten minimieren. Ich gehe hier nur auf die Wege ein, die in meiner Arbeit theoretisch möglich sind:

- (1) Homogenisieren
- (2) Ausbalancieren

Um diese beiden Wege zu erläutern, greife ich etwas vor und führe die Variable *Geschlecht* als potenzielle SV ein. Demnach kann es sein, dass die Wahl der sprachlichen Mittel in *GEolino* und *GEO* vom Geschlecht des jeweiligen Autors abhängt. Der Einfluss der SV lässt sich einerseits dadurch minimieren, dass man nur entweder Texte von männlichen oder Texte von weiblichen Autoren in das Korpus aufnimmt. Damit wäre die SV homogenisiert. Andererseits kann man die SV ausbalancieren, indem man die Verteilung der Texte von männlichen und weiblichen Autoren gleichmäßig gestaltet, das heißt in beide Korpora gleich viele Texte von männlichen und weiblichen Autoren aufnimmt. Im Zusammenhang mit den Kriterien (2) bis (4) entstehen drei Störvariablen, die ich genauer erläutere.

4.1.1 Themengleichheit

Nach Kriterium (2) nehme ich nur Texte in die Korpora auf, die den gleichen Themenbereich behandeln. Nach Rosebrock/Nix (2008) sind Sachtexte semantisch domänenspezifisch. Ich gehe davon aus, dass sich dies auch sprachlich in den Texten widerspiegeln kann. So wird zum Beispiel ein Text über Dinosaurier allein durch die Schreibung der vollständigen Namen potenziell längere Wörter und vielleicht auch mehr Fachbegriffe enthalten als ein Text über das Zirkusleben. Vergleiche ich nun diese beiden Texte bezüglich der Merkmale *Wortlänge* und *Fachbegriffe* quantitativ miteinander, kann ich diesbezügliche Unterschiede nicht eindeutig auf die UV *Korpus* zurückführen. Die SV *Thema* kann ich dadurch homogenisieren, dass ich nur Texte zu einem einzigen Themenbereich in die Korpora aufnehme. Dadurch wird der Einfluss der SV minimiert.

4.1.2 Metadaten

Ich habe oben bereits erläutert, wie man die SV *Geschlecht* homogenisieren bzw. ausbalancieren kann. Beide Wege sorgen dafür, dass der Einfluss der SV minimiert würde.

Bei der Zusammenstellung der Texte für die Korpora wird jedoch schnell klar, dass ich diese SV nicht ausbalancieren kann und somit Kriterium (3) nicht erfüllbar ist.

Dies liegt an einem grundsätzlichen Ungleichgewicht in der Geschlechterverteilung in den beiden Zeitschriften: Für die *GEolino* schreiben ungleich viel mehr Frauen als Männer; von elf Texten wurde lediglich ein Text von einem Mann verfasst. Für die *GEO* gilt das Gegenteil: Von sieben Texten wurden nur zwei Texte von Frauen geschrieben. Es war mir nicht möglich, sowohl eine gleichmäßige Geschlechterverteilung einzuhalten als auch annähernd gleich lange Texte auszuwählen (vgl. Kriterium (1)). Aus dem gleichen Grund ist die SV auch nicht homogenisierbar.

Ein weiteres Problem stellen Texte dar, die von mehreren Autoren verfasst wurden. Bei diesen Texten stellt sich generell die Frage, welchem Geschlecht man sie zuschreibt. Diese Überlegung ist nötig, weil sowohl *GEolino7* als auch *GEO40* solche Gemeinschaftstexte sind. *GEolino7* wurde von zwei Autoren verfasst, *GEO40* von vier Autoren. *GEolino7*

ließe sich noch problemlos dem Geschlecht *w* zuordnen, da beide Autoren weiblich sind. Die Autorengruppe von *GEO40* setzt sich jedoch sowohl aus weiblichen als auch aus männlichen Autoren zusammen. Ich habe mich dafür entschieden, Gemeinschaftstexte dem Geschlecht zuzuschreiben, dessen Autoren quantitativ überwiegen. Da die Autorengruppe von *GEO40* aus drei Männern und einer Frau besteht, schreibe ich diesen Text also dem Geschlecht *m* zu.

Ein weiteres Problem sind anonyme Texte. Die Autoren der Texte *GEOLino14* und *GEOLino20* waren auch nach ausführlicher Recherche des Verlags *Gruner und Jahr* nicht mehr zu ermitteln. Diese Texte habe ich somit keinem Geschlecht zugeordnet.

Da das Verhältnis von männlichen und weiblichen Autoren mit 1:8 für die *GEOLino* und 5:2 für die *GEO* so unausgeglich ist, liegt die Vermutung nah, dass eventuelle Unterschiede zwischen den Korpora nicht der *UV Korpus* zugeschrieben werden können, sondern durch die *SV Geschlecht* bedingt sind. Ob das Geschlecht eines Autors tatsächlich einen Einfluss darauf hat, wie der Text sprachlich gestaltet ist, prüfe ich in Abschnitt 5.3.1 durch eine Voranalyse.

4.1.3 Autoren

Für jedes Korpus gilt nach Kriterium (4): Alle Texte stammen von verschiedenen Autoren. Das heißt, dass weder im Korpus *GEOLino* noch im Korpus *GEO* zwei Texte vom selben Autor stammen. Wenn ich mehrere Texte vom selben Autor in ein Korpus aufnehmen würde, könnten Unterschiede zwischen den Korpora darin begründet sein, dass ein bestimmter Autor auf eine spezifische Art und Weise schreibt. Somit wäre der Einfluss der *UV Korpus* auf die *AV* nicht mehr eindeutig, da die *SV Autor* ebenfalls einen Einfluss hat. Durch Kriterium (4) versuche ich, diese *SV* auszubalancieren.

Jedoch schreiben zwei der Autoren sowohl für die *GEOLino* als auch für die *GEO*. Daher habe ich jeweils einen Text von ihnen in beide Korpora aufgenommen. Innerhalb der Korpora ist also nach wie vor jeder Autor nur einmal vertreten.

Ob die *SV Autor* tatsächlich ausbalanciert ist, prüfe ich ebenfalls durch eine Voranalyse. Wenn sich herausstellt, dass die *SV Autor* nicht ausbalanciert werden konnte, erübrigt sich die Voranalyse im Rahmen der *SV Geschlecht*: Wenn schon innerhalb einer Geschlechtergruppe signifikante Unterschiede existieren, bedeutet das, dass es die *SV Autor* gibt und diese ist stärker als die *SV Geschlecht* (vgl. Abschnitt 5.3.1) (Quelle: Felix Golcher, persönliches Gespräch).

4.1.4 Metasprachliche Elemente

Ich habe alle textstrukturierenden Elemente – Überschriften, Unterüberschriften und Teaser – in den Korpora beibehalten, damit die Texte vollständig bleiben und nicht verfälscht werden. Für meine Arbeit haben diese Elemente keine weitere Bedeutung (vgl. Abschnitt 4.2.13).

4.1.5 Eingriffe

Ich habe die Texte korrigiert, da sich folgende Fehlertypen ergaben:

- (1) Orthografiefehler
- (2) fehlende Spatien
- (3) Abfolgefehler
- (4) formale Fehler im Worttrennungsbereich

Die Fehlertypen (3) und (4) möchte ich kurz näher erläutern. In einem der Texte sind Phrasen eines Textblocks durcheinander geraten, so dass dieser nicht mehr sinnvoll zu lesen war. Damit dies nicht zu Verfälschungen (insbesondere bezüglich der Satzlänge) führt, habe ich die Abfolge korrigiert.

In den elektronischen Texten wurde keine automatische Silbentrennung benutzt. Stattdessen wurden die Wörter manuell getrennt. Da ich die Texte in ein Textverarbeitungsprogramm übernahm, um weiter zu verarbeiten, veränderte sich die Zeilenbreite. Aufgrund der manuellen Worttrennung traten dann Fälle wie *ge-macht* auf. Diese formalen Fehler habe ich entsprechend korrigiert.

4.1.6 Erstellungswerkzeug

Ich habe die Korpora mithilfe des Treetaggers erstellt¹⁴. Durch den Treetagger ist eine dreifache Vorverarbeitung der Texte möglich: Die Texte werden automatisch tokenisiert, das heißt Wortgrenzen werden automatisch bestimmt. Außerdem wird eine Lemmatisierung der Texte vorgenommen, durch welche die Grundform jedes Wortes annotiert wird. Zudem wird die Wortart jedes Wortes durch die *part-of-speech*-Ebene (POS) vorannotiert¹⁵.

4.1.7 Zusammensetzung der Korpora

Hunston (2008:156) geht davon aus, dass „corpora are a compromise between what is desirable [...] and what is possible“. Korpora stellen also einen Kompromiss zwischen dem Forschungsziel und der Realität dar. Auch eine präzise Fragestellung garantiert kein ideales Korpus (ebd.). Daher lassen sich auch die von mir zusammengestellten Korpora sicherlich noch optimieren (mehr dazu im Abschnitt 7.1).

Meine Korpora setzen sich folgendermaßen zusammen: Sie bestehen jeweils aus ca. 10.000 Token und beinhalten verschiedene Subkorpora (die einzelnen Texte). Die Anzahl der Subkorpora unterscheidet sich dabei stark. Während das Korpus *GEolino* aus sieben Subkorpora besteht, enthält das Korpus *GEO* elf Subkorpora. Bereits hier wird deutlich, dass die Texte in der *GEolino* im Schnitt kürzer sind als in der *GEO*, da beide Korpora trotz der unterschiedlichen Textanzahl ungefähr gleich viele Token aufweisen.

Die Metadaten *Name* und *Geschlecht* der einzelnen Autoren sowie die Zuordnung zu den Texten habe ich in separaten Tabellenblättern festgehalten.

¹⁴<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/> vom 12.3.12.

¹⁵<http://www.ims.uni-stuttgart.de/projekte/corplex/TagSets/stts-table.html> vom 12.3.12.

Tabelle 1 stellt eine Übersicht der Korpora dar:

	<i>GEolino</i>	<i>GEO</i>
Anzahl Token	10479	10677
Anzahl Texte	11	7
Gemeinschaftstexte	1 (2 Autoren)	1 (4 Autoren)
Anzahl Texte _m	1	5
Anzahl Texte _w	8	2
Anzahl anonyme Texte	2	0

Tabelle 1: Korpusübersicht

4.2 Annotation und Richtlinien

Ich habe meine Korpora in Microsoft Excel annotiert. Dabei sind einige Merkmale im Konjunktions- und Adjektivbereich durch die POS-Ebene automatisch vorannotiert. Für die restlichen Merkmale war es nötig, möglichst genaue Annotationsrichtlinien festzulegen. In diesen Richtlinien halte ich fest, welche Elemente ich unter welchen Voraussetzungen als bestimmte sprachliche Merkmale annotiere und unter welchen Umständen ich sie nicht annotiere. Nur so ist eine konsistente Annotation der Korpora möglich.

Die von mir ausgewählten Merkmale sind fast vollständig dem Bereich *Komplexität* zuzuordnen und zwar im Sinne von Erweiterungen auf verschiedenen sprachlichen Ebenen. Lediglich ein Merkmal lässt sich nicht in diesen Bereich einordnen (mehr dazu in Abschnitt 4.2.12). In den folgenden Abschnitten erläutere ich meine Richtlinien ausführlich. Dazu gehe ich noch einmal kurz auf die Motivation des jeweiligen Merkmals ein, die sich aus den Ausführungen im Abschnitt 3.2 ergibt. Gegensätzliche Annahmen zum gleichen Merkmal sind für mich besonders interessant, da ich hier die Möglichkeit zur empirischen Überprüfung habe und den Gegensatz im kleinen Rahmen entscheiden kann. Ich stelle die tagsets der Merkmale vor und formuliere Erwartungen, die sich aufgrund der Motivationen ergeben. Alle Erwartungen ergehen unter der Voraussetzung, dass die *GEolino* ihren Lesern sprachlich „entgegenkommt“: Demnach ist sie bemüht, ihren Lesern gut verständliche, informative und unterhaltsame Texte anzubieten.

Eine Übersicht der Richtlinien stelle ich im Anhang A zusammen. Für alle Merkmale gilt, dass ich rein fremdsprachliches Material nicht annotiere. Zitate behandle ich als normalen Text und annotiere sie ebenfalls. Bezüglich des Merkmals *Leserkontakt* bilden Zitate jedoch einen Sonderfall (siehe Abschnitt 4.2.12).

4.2.1 Oberflächenmerkmale

Aus allen von mir betrachteten Bereichen geht hervor, dass die Wort- und Satzlänge die Lesbarkeit und das Verstehen eines Textes beeinflusst. Zudem nimmt man in den Bereichen *Journalismus* (Kommerell (2008)), *Sachtextforschung/Textlinguistik* (Baumann (2009)) und *Lesbarkeitsforschung* (Groeben (1982)) an, dass kurze Wörter und Sätze vorgezogen werden sollten.

Methode

Mit dem Statistik-Programm *R* habe ich die Ebene *WL* für die Wortlänge direkt in die Korpora eingefügt. Für jedes Wort in jedem Subkorpus habe ich die Länge in Zeichen

automatisch errechnet. Dabei habe ich als Untergrenze für die Wortlänge zwei Zeichen festgelegt. So stelle ich sicher, dass Satzzeichen nicht mitgezählt werden. Diese sind rein formal einbuchstabige Wörter und werden von *R* auch als solche interpretiert. Die Untergrenze von zwei Zeichen ist nötig, weil die Satzzeichen sonst die Ergebnisse verfälschen würden. Weitere Einschränkungen treffe ich hier nicht. Man könnte einen Großteil der Funktionswörter ausschließen, indem man die Untergrenze auf zum Beispiel vier oder fünf Zeichen anhebt. Auf diesem Wege würde man mit einer höheren Wahrscheinlichkeit lediglich Inhaltswörter betrachten (Quelle: Felix Golcher, persönliches Gespräch). Geht man von der Annahme aus, dass eher im Bereich der Inhaltswörter Varianz auftreten kann, wäre dieser Weg sinnvoll. Ich gehe hier jedoch zunächst davon aus, dass auch im Bereich der Funktionswörter genug Varianz auftreten kann, so dass signifikante Unterschiede entstehen können.

Auch die Satzlänge der einzelnen Subkorpora habe ich mit *R* direkt errechnet. Dazu habe ich das Verhältnis von Tokenanzahl zur Anzahl der Satzzeichen ermittelt.

Erwartung

Folglich den Annahmen von Kommerell (2008), Baurmann (2009) und Groeben (1982), sollten die Texte in der *GEOLino* eine geringere Wort- und Satzlänge aufweisen als die Texte in der *GEO*. Auch die Annahmen bezüglich der Komposita (siehe Abschnitt 4.2.2) legen die Erwartung nahe, dass die Wörter in der *GEOLino* kürzer sind als in der *GEO*.

4.2.2 Wortbildung: Komposition

Aus dem Bereich der Sachtextforschung/Textlinguistik (Gierlich (2005)) kommt der Hinweis, dass sich ein Text, in dem Komposita hochfrequent verwendet werden, negativ auf das Verstehen auswirkt. Durch Komposition werden die betreffenden Wörter per se länger, daher greift hier zudem die gleiche Motivationsgrundlage wie für das Oberflächenmerkmal Wortlänge. Ich beschränke mich auf Komposita im Substantivbereich, da diese den Großteil der Komposita im Deutschen ausmachen (vgl. Eisenberg (2004:226)).

Methode

Ich annotiere auf der Ebene *nominales.kompositum* alle rechtsköpfigen Komposita mit nominalem Kopf als *nk*.¹⁶ Dabei kann das Kompositum auch als Bindestrich-Kompositum auftreten. Ich richte mich dabei nach den *POS-tagsNN* und *TRUNC*. Das Erstglied kann in Form eines Substantivs (*Raumanzug*), Adjektivs (*Echtzeit*), Verbs (*Messwert*), oder einer Präposition (*Außenkamera*) auftreten (vgl. Eisenberg (2004:226)). Bei zusammengefassten Komposita, die sich über die Token-Grenze hinweg erstrecken (*Bild- und Messdaten*), trägt das Erstglied den *POS-tagTRUNC*, das Zweitglied den *POS-tagNN*. Ich annotiere jedes Glied mit *nk*.

Eigennamen (*POS-tagNE*) betrachte ich als Komponente der Komposition (*Apollo-8-Astronaut*), sofern die Komposition nicht ausschließlich aus dem Eigennamen besteht, wie *Raumfahrtbehörde* in *Europäische Raumfahrtbehörde* (ESA).

Nicht alle Komposita fallen gleichermaßen als solche auf. Sicher wird man *Raumanzug* eher als Kompositum wahrnehmen als zum Beispiel *Rucksack* oder *Bahnhof*. Letztere sind

¹⁶ Nach Bußmann (2002:469) verwende ich die Begriffe *Nomen* und *Substantiv* synonym.

lexikalisierte Komposita, die nicht mehr als Zusammensetzung, sondern als lexikalische Einheit wahrgenommen werden.

Nach Eisenberg (2004:215) gehören solche Wörter zum Wortschatz und sind im Lexikon als usuelle Bildung gespeichert (vgl. auch Bußmann (2002:405)). Lexikalisierung ist nicht kategorisch, sondern ein Prozess, der (auch subjektive) Grauzonen zulässt. Dasselbe Wort kann von mehreren Menschen sowohl als lexikalisiert als auch als nicht lexikalisiert aufgefasst werden. Daher gibt es keine Listen oder formalen Merkmale, nach denen man zuverlässig und eindeutig bestimmen kann, ob ein Wort lexikalisiert ist oder nicht. Ob ein Kompositum lexikalisiert ist oder nicht, liegt hier also in meinem Ermessen. Komposita, die ich als lexikalisiert deklariere, interpretiere ich als Simplizia und annotiere sie folglich nicht als Komposita. Auf diese Weise ist es mir möglich, tatsächlich nur die „echten“ Komposita zu betrachten. Eine Liste der lexikalisierten Komposita füge ich im Anhang B bei.

Erwartung

Folglich den Annahmen von Gierlich (2005) sowie den Annahmen zum Oberflächenmerkmal *Wortlänge*, müssten die Daten einen Unterschied zwischen den Texten der *GEolino* und der *GEO* zeigen: Ich erwarte in der *GEolino* eine niedrigere Frequenz von nominalen Komposita als in der *GEO*.

4.2.3 Anzahl der Hauptglieder

Nominale Komposita sind im Deutschen äußerst produktiv und theoretisch unendlich erweiterbar. Auch hier greift die gleiche Motivationsgrundlage wie für das Oberflächenmerkmal *Wortlänge*: Jedes Kompositionsglied verlängert das Wort und Wortlänge wird in vielen Arbeiten als relevanter Parameter für das Verstehen eines Textes angesehen.

Methode

Auf der Ebene *anzahl.glieder* annotiere ich freie Morpheme, die in einem Kompositum auftreten, mit numerischen *tags*. Da ein Kompositum aus mehreren, in sich wieder komplexen Gliedern bestehen kann, nehme ich hier eine Einschränkung vor: Ich betrachte nur die Kompositionsglieder auf der hierarchisch höchsten Ebene, also die Hauptglieder des Kompositums.

Erwartung

Folglich den Annahmen zum Oberflächenmerkmal *Wortlänge*, dürfte auch hier ein Unterschied zwischen *GEolino* und *GEO* zu erwarten sein: In der Absicht, Wörter möglichst kurz zu halten, würde die *GEolino* tendenziell Komposita verwenden, die nicht mehr als zwei Hauptglieder aufweisen. In der *GEO* hingegen würde ich eher Komposita erwarten, die aus mehreren Hauptgliedern bestehen. Die Einschränkung auf die Hauptglieder lässt jedoch noch eine weitere Erwartung zu: Die meisten Komposita im Deutschen sind zweigliedrig (vgl. Klosa/Scholze-Stubenrecht/Wermke(1998:411)). Daher wäre auch zu erwarten, dass sich *GEolino* und *GEO* in diesem Punkt nicht signifikant voneinander unterscheiden.

4.2.4 Komplexität der Hauptglieder

Wortbildungs(teil)verfahren sind hierarchisiert, das heißt dass ein Kompositum aus zwei Hauptgliedern bestehen kann, diese wiederum theoretisch aus hunderten Gliedern. Das Kompositum *Funkverbindung* besteht zum Beispiel aus zwei Hauptgliedern, diese sind wiederum Ergebnisse anderer Wortbildungsverfahren (Derivation und Konversion), so dass das Kompositum letztendlich aus den vier Elementen *Funk+ver+bind+ung* besteht. Während das Kompositum also auf der obersten Ebene einfach aufgebaut ist, weisen die Glieder auf tieferen Ebenen durchaus Komplexität auf. Deswegen wäre es nicht aussagekräftig genug, Komposita nur auf der obersten Ebene zu betrachten.

Methoden

Auf der Ebene *komplexitaet.glieder* annotiere ich tiefsitzende Komplexität mit *ko* für mindestens ein komplexes Glied und *si* für ausschließlich simplexe Glieder. Ich beziehe hier neben der Komposition Derivation und Konversion als mögliche Wortbildungsverfahren ein. Unter Derivation verstehe ich nach Bußmann (2002:155) und Pelz (2004:128f) Ableitungen (Präfix- und Suffixbildungen), die Kategorienverändernd sein können, aber nicht müssen (vgl. Eisenberg (2004:272,274)). So wirkt zum Beispiel das Suffix *-keit* in *Einsamkeit* Kategorienverändernd, indem es ein Adjektiv in ein Substantiv umwandelt. Im Fall von *Menschheit* wirkt das Suffix *-heit* Kategorienverhaltend.

Konversion verstehe ich nach Pelz (2004:129f) als „das Überwechseln eines Wortes in eine andere Wortart ohne Veränderung seiner Ausdrucksseite“, also als eine Ableitung durch ein Nullmorphem (ebd.). Entgegen Eisenberg (2004:295) und in Anlehnung an Bußmann (2002:380) zähle ich auch die Fälle zu den Konversionen, bei denen der Kategorienwechsel durch einen Vokalwechsel erreicht wird, wie zum Beispiel bei *abfliegen* – *Abflug*. Ich folge Eisenbergs (2004:295) Annahme, dass Konversionen in einer bestimmten Reihenfolge ablaufen. Dabei ist der letzte Schritt einer Konversion der Wandel vom Verb zum Substantiv. Daher wird sofort deutlich, dass zum Beispiel *flug* in *Abflug* ein komplexes Hauptelement ist. Hier liegt das Verb *fliegen* zugrunde, also annotiere ich *Abflug* als *ko*. Alle Kompositions-Hauptglieder, die nach den oben genannten Definitionen weder Komposition noch Derivation oder Konversion beinhalten, interpretiere ich als Simplicia und annotiere sie als *si*.

Ich möchte die Richtlinie am Beispiel *Funkverbindung* veranschaulichen. Ich habe oben bereits erwähnt, dass sich dieses Kompositum aus den zwei Hauptgliedern *Funk* und *Verbindung* zusammensetzt, die in sich komplex sind. Die Komponente *Funk* ist komplex, weil ich die Konversionsrichtung Verb > Substantiv annehme. Die Komponente *Verbindung* ist ebenfalls komplex, da hier Derivation in Form des Präfixes *ver-* und des Suffixes *-ung* vorliegt. Daher annotiere ich dieses Kompositum als *ko*.

Ein einfaches Beispiel ist *Raumschiff*. Dieses Kompositum besteht aus zwei Hauptgliedern, die beide Simplicia sind. Daher annotiere ich das Kompositum hier als *si*.

Fußballfeld hingegen ist ein Kompositum, beidem das erste Hauptglied *Fußball* selbst ein Kompositum und daher komplex ist, während das zweite Hauptglied *Feld* ein Simplex ist. In diesem Fall annotiere ich das Kompositum *Fußballfeld* aufgrund der Komplexität des ersten Glieds als *ko*.

Erwartung

Aus den Annahmen zum Oberflächenmerkmal Wortlänge und aus Kommerells Annahme, kurze Formulierungen seien für junge Leser gut geeignet, weil so der kognitive Arbeitsaufwand reduziert wird (2008:132) folgt: In der *GEOlino* erwarte ich weniger komplexe Kompositionsglieder als in der *GEO*.

4.2.5 Adjektivfrequenz

Aus den Bereichen der Lesbarkeitsforschung (May (1996)) und des Journalismus (Kommerell (2008)) kommen gegensätzliche Hinweise zum Gebrauch von Adjektiven. Während May (1996) den hochfrequenten Gebrauch von Adjektiven positiv bewertet, rät Kommerell (2008) dringend davon ab. Diesen Gegensatz möchte ich im Rahmen von *GEOlino* und *GEO* empirisch überprüfen.

Methode

Die Annotation von prädikativen und attributiven Adjektiven ist automatisch durch die POS-Ebene gegeben (POS-tags: *ADJA* und *ADJD*). Daher ist kein separates Tagset notwendig.

Erwartung

Die Formulierung einer Erwartung ist schwierig, denn sie hängt davon ab, welchem Ansatz man folgt. Nach May (1996) erwarte ich in der *GEOlino* eine höhere Adjektivfrequenz als in der *GEO*. Nach Kommerell (2008) hingegen erwarte ich das genaue Gegenteil.

4.2.6 Attributive und prädikative Adjektive

Aus dem Bereich der Psycholinguistik (Doderer (1960)) kommt die Annahme, dass die stärkere Verwendung von attributiven Adjektiven eine besondere Rolle in der Sprachentwicklung von Kindern spielt. Aus dem Bereich der Registerforschung (Biber (1995)) kommt zudem der Hinweis, dass die Verwendungshäufigkeit von attributiven und prädikativen Adjektiven Register unterscheidend sein kann. Auch um einen Schritt weiter in Richtung der Fragestellung (2) zu gehen, prüfe ich diese Möglichkeit des Unterschiedes hier empirisch.

Zur Methode siehe Abschnitt 4.2.5.

Erwartung

Folglich Doderer (1960) und Biber (1995), müssten sich Texte aus der *GEOlino* von denen der *GEO* bezüglich der Verwendung von attributiven und prädikativen Adjektiven unterscheiden. Nach Doderer (1960) würde ich mehr attributive Adjektive in der *GEOlino* erwarten als in der *GEO*.

4.2.7 Komparation

Wie schon bezüglich der Adjektivfrequenz gibt es im Hinblick auf den Gebrauch von Komparation gegensätzliche Annahmen in der Literatur: Aus dem Bereich des Journalismus (Kommerell (2008)) geht hervor, dass Superlative unbedingt zu vermeiden sind. Im Bereich der Textlinguistik/Sachtextforschung (Gierlich (2005)) gibt es hingegen die Annahme, dass der Einsatz von Komparation eine wichtige Rolle in der

Wissensvermittlung spielt und daher zu empfehlen ist. Diesen Gegensatz möchte ich im Rahmen von *GEolino* und *GEO* empirisch überprüfen.

Methode

Ich annotiere auf der Ebene *komparation* alle drei Komparationsstufen der Adjektive: *pos* für den Positiv, *komp* für den Komparativ und *sup* für den Superlativ. Dabei orientiere ich mich an den POS-tags *ADJA*, *ADJD* und *TRUNC* und annotiere einerseits Adjektive, die als Substantiv-Erweiterungen fungieren und andererseits Adjektive, die in Form von Adverbien als Verberweiterungen fungieren.

Homographen, die durch Flexion entstehen, wie zum Beispiel *ein gefährlicherer Traum*, tragen formal die Endung eines Komparativs. Diese Fälle annotiere ich jedoch nicht als *komp*, sondern als *pos*.

Adverbien wie *weitere (Monate)* oder *spätestens* annotiere ich nicht, da diese in der Form eine spezielle Bedeutung haben und nur als formaler Komparativ bzw. Superlativ auftreten können, wenn der Satz grammatisch sein soll, vgl.:

- (1) *Sie blieben weitere drei Monate oben.* vs. **Sie blieben weite/weiteste drei Monate oben.*
- (2) *Sie kommen spätestens in drei Monaten wieder.* vs. **Sie kommen spät/später in drei Monaten wieder.*

Ein wichtiger Punkt sind Absolutadjektive. Diese Adjektive sind nicht komparierbar, somit hat ein Autor hier keine Wahl und muss den Positiv verwenden, vgl.:

- (3) *dreieckig* vs. **dreieckiger/*am dreieckigsten*
- (4) *blau* vs. **blauer/*am blausten*.

Hierzu zähle ich auch die Zahladjektive. Ich annotiere Absolutadjektive nicht, weil sie einen eventuellen Komparationseffekt in den Daten verfälschen würden. Ebenso annotiere ich Adjektive in festen Wendungen bzw. Eigennamen nicht, so zum Beispiel *Kalter Krieg*, *unbestimmter Artikel*, *danke schön*, *so gut wie*, *Großer Wagen* oder *rechter Winkel*.

Partizipien wie *glitzernd* oder *erschwert* annotiere ich nicht. Sind sie jedoch als Adjektive lexikalisiert, wie zum Beispiel *verrückt*, behandle ich sie wie normal steigerbare Adjektive und annotiere sie entsprechend ihrer Komparationsstufe als *pos*, *komp* oder *sup*.

Scheinbare Adjektive, die nur aufgrund von trennbaren Verben entstehen, wie *frei insetzt Energie frei*, annotiere ich nicht, da sie keine echten Adjektive sind.

Erwartung

Folglich den Annahmen von Kommerell (2008), würde ich in der *GEolino* eine geringere Frequenz von Superlativen erwarten als in der *GEO*. Nach Gierlich (2005) hingegen müsste die *GEolino* generell mehr Gebrauch von Komparation aufweisen als die *GEO*.

4.2.8 Adjektivische Amplifikation

Adjektive wie *blitzblank*, *nagelneu* oder *haarscharf* müssen laut Richtlinie zur Komparation von der Annotation ausgeschlossen werden. Diese Adjektive sind jedoch weder normale Adjektive in dem Sinne, dass sie komparierbar sind, noch sind sie reine Absolutadjektive wie zum Beispiel *blau* oder *dreieckig*. Sie beinhalten vielmehr wortintern

eine – wenn nicht zwei – Komparationsstufe(n) in Form eines (manchmal nicht mehr transparenten) Vergleichs-Präfixoids bzw. -Präfixes (vgl. Bußmann (2002:105)).¹⁷ Daher weisen sie schon gewisse Merkmale von Komparation auf und dürfen daher nicht unbeachtet bleiben.

Methode

Ich annotiere auf der Ebene *amplifikation* nicht steigerbare Adjektive, die Vergleichs-Präfix(oid)e aufweisen, als *ampl.* Ich orientiere mich hier an den POS-tags *ADJD*, *ADJA* und *TRUNC*.

Dabei annotiere ich auch die Adjektive, deren Präfix(oid)e keine eindeutige Steigerung anzeigen, wie es bei den oben genannten Beispielen der Fall ist. Daher zähle ich auch *meterhoch*, *stundenlang* oder *turmhoch* zu den adjektivischen Amplifikationen. Nur wenn zwar ein wortinhärenter Vergleich vorliegt, jedoch eindeutig keine Steigerung angezeigt wird, annotiere ich das Adjektiv nicht als *ampl.*, zum Beispiel bei *schubladengroß* oder *bogenförmig*.

Alle Adjektive, die nicht amplifiziert wurden, annotiere ich als *noampl.*

Erwartung

Hier lassen sich zwei Erwartungen formulieren. In den adjektivischen Amplifikationen stecken Komparationsstufen, daher würde ich die erste Erwartung analog zur „normalen“ Komparation sehen: Nach Kommerell (2008) müssten adjektivische Amplifikationen in der *GEOlino* seltener auftreten als in der *GEO*; folglich Gierlich (2005), würde ich genau das Gegenteil erwarten.

Adjektivische Amplifikationen sind sehr anschauliche Wörter, allein durch den wortinhärenten Vergleich. Da nach Kommerell (2008:140ff., vgl. auch Niederhauser (1999:213), Gierlich (2005:29), Baurmann (2009:50), Kühn (2010:132), Gibson/Levin (1989:269f)) junge Leser ein hohes Maß an Anschaulichkeit stärker brauchen als Erwachsene, um einen Text zu verstehen, lässt sich die zweite Erwartung in Anlehnung an Gierlich (2005) formulieren: Die Texte in der *GEOlino* weisen mehr adjektivische Amplifikationen auf als die Texte in der *GEO*.

4.2.9 Nominale Abstrakta

Arbeiten aus den Bereichen der Lesbarkeitsforschung (DuBay (2004), Groeben (1982)), Sachtextforschung/Textlinguistik (Baurmann (2009), Christmann/Groeben (2006)) und des Journalismus (Kommerell (2008)) legen nahe, dass der Gebrauch von abstrakten Begriffen das Textverständnis beeinflusst. Fast durchgängig geht man davon aus, dass abstrakte Begriffe für junge Leser ungeeignet sind, da sie aus semantischen Gründen das Verständnis hemmen und somit kognitiv zu komplex sind.

Methode

Ich beschränke mich hier auf Abstrakta im Substantivbereich und annotiere auf der Ebene *nominales.abstraktum* abstrakte Nomen als *abst.* Annotationsgrundlage bildet eine Liste von Groeben (1982:226) – in verkürzter Form auch bei Kommerell (2008:130). Laut dieser Liste zeigen die Suffixe *-heit*, *-ie* [i:], *-ik*, *-ion*, *-ismus*, *-ität*, *-keit*, *-nz*, *-tur* und -

¹⁷ Eine ganz ähnliche Konstruktion findet man in der lateinischen Komparation in Form des Elativs.

ungssubstantivische Abstraktheit an. Ich richte mich beim Annotieren nach den POS-tags *NN* und *TRUNC*. Alle anderen Substantive (mit Ausnahme der Eigennamen) annotiere ich als *noabst*. Offensichtlich nicht abstrakte Substantive wie *Spion*, *Natur*, *Sprung* oder *Tanz* annotiere ich als *noabst*.

Nominale Komposita annotiere ich nur als *abst*, wenn das Zweitglied – der Kopf – des Kompositums eines der relevanten Suffixe aufweist, wie zum Beispiel *Spiralgalaxie*. Komposita, bei denen lediglich das Erstglied eines der relevanten Suffixe aufweist, annotiere ich nicht als *abst*, denn hier ist nur ein Teil des Kompositums per definitionem abstrakt, nicht jedoch das ganze Kompositum. Der Kopf ist ausschlaggebend dafür, ob das Wort als Ganzes als abstrakt interpretiert wird oder nicht. So ist eine *Galaxie* zwar nach Groebens Liste ein Abstraktum, der *Galaxienhaufen* jedoch nicht. Abstrakte Komposita, die über die Token-Grenze (angezeigt durch |) hinweg gehen, annotiere ich am Zweitglied, zum Beispiel *D2-Mission*.

Ich gehe davon aus, dass Eigennamen wie *Internationale Raumstation* oder *Berliner Philharmonie* einen hohen Lexikalisierungsstatus haben. Der Autor hat bei Eigennamen auch keine Ausdrucksalternative. Daher annotiere ich Eigennamen nicht als *abst*, auch wenn sie eines der relevanten Suffixe aufweisen.

Erwartung

Folglich den Annahmen aus der Lesbarkeitsforschung, der Sachtextforschung/Textlinguistik und dem Journalismus, würde ich erwarten, dass die Texte in der *GEOlino* weniger Abstrakta aufweisen als die Texte in der *GEO*.

4.2.10 Konjunktionsfrequenz

Auch zur allgemeinen Frequenz von Konjunktionen gibt es gegensätzliche Annahmen im Bereich Lesbarkeitsforschung: Während von Willenberg (2005) der Hinweis kommt, dass Konjunktionen generell Verständnis fördernd wirken, nimmt Groeben (1982) Konjunktionen als potenziell Verständnis hemmend wahr.

Methode

Konjunktionen sind automatisch als *KON*, *KOUI*, *KOUS* und *KOKOM* auf der POS-Ebene vorannotiert. Daher entfällt ein separates tagset hier.

Erwartung

Nach Willenberg (2005) würde ich in der *GEOlino* einen stärkeren Gebrauch von Konjunktionen erwarten als in der *GEO*. Folglich jedoch Groeben (1982), müsste genau das Gegenteil der Fall sein: In der *GEOlino* würden weniger Konjunktionen auftreten als in der *GEO*.

4.2.11 Semantische Klassen von Konjunktionen

Adversative und kausale Konjunktionen verleihen einem Text vielleicht ein höheres Maß an Explizität: Sie zeigen dem Leser an, wie er die Sachverhalte zueinander in Beziehung setzen muss und wie sie zu verstehen sind. Aus dem Bereich der Sachtextforschung geht in diesem Zusammenhang hervor, dass adversative und kausale Konjunktionen eine Verständnis fördernde Wirkung haben, da sie einem Sachtext Kohärenz verleihen (vgl. Hiller (2010)). Aus dem Bereich der Psycholinguistik (Feilke (1996)) kommt jedoch der

Hinweis, dass diese Explizitt fr junge Leser nicht wichtiger ist als fr Erwachsene. Feilke (1996) weist zudem darauf hin, dass Kinder kausale Konjunktionen ab einem Alter von ca. 10 Jahren auch selbst deutlich weniger gebrauchen als zuvor. Daher kann man annehmen, dass sie auch bei der Sprachrezeption nicht mehr auf eine eventuell vorhandene Verstndnisfrdernde Wirkung dieser Konjunktionen angewiesen sind.

Methode

Ich annotiere semantische Klassen von Konjunktionen auf der Ebene *konjunktion.semantisch*. Dabei beschrnke ich mich auf adversative und kausale Konjunktionen, um die Annahmen von Hiller (2010) und Feilke (1996) zu prfen. Bei der Annotation richte ich mich nach der Liste in Buscha (1989). Das heit, dass ich mich hier nicht an der POS-Ebene orientiere, sondern nach bestimmten Konjunktionen suche und diese als *advers* oder *kaus* annotiere.

Erwartung

Folglich Hiller (2010), wrde ich in der *GEOLino* mehr adversative und kausale Konjunktionen erwarten als in der *GEO*, denn man wrde diese expliziten sprachlichen Mittel als Verstndnishilfe strker fr junge Leser einsetzen. Nach Feilke (1996) wrde ich jedoch keinen signifikanten Unterschied im Gebrauch von adversativen und kausalen Konjunktionen erwarten. Insbesondere die Anzahl kausaler Konjunktionen drfte sich in der *GEOLino* nicht von der *GEO* unterscheiden, da Feilke (1996) gezeigt hat, dass die Kinder selber diese Klasse von Konjunktionen schon ab einem Alter von 10 Jahren immer weniger gebrauchen.

4.2.12 Leserkontakt

Aus den Bereichen der Sachtextforschung/Textlinguistik (Baurmann (2009)), Textverstndlichkeitsmodelle (Gpferich (2008), Schulz/Tausch/Langer (2011), May (1996)) und des Journalismus (Kommerell (2008)) geht einstimmig hervor, dass die direkte Anrede an den Leser besonders fr junge Leser einen positiven Einfluss auf die Lesemotivation und das Verstndnis hat. Aus dem Bereich der Registerforschung (Biber/Conrad/Reppen (2002), Biber (2006)) kommt auerdem der Hinweis, dass die Verwendung von Personalpronomen der 2. Person in geschriebenen Informationstexten fr Erwachsene nicht blich ist. Daher hat die Verwendung dieser Pronomen eine Register unterscheidende Funktion (Biber/Conrad (2009)).

Methode

Ich annotiere Leserkontakt auf der Ebene *anrede* als *anr*. Analog zu Biber/Conrad/Reppen (2002), Biber (2006) und Biber/Conrad (2009) interpretiere ich Personalpronomen in der 2. Person (*du/Ihr, Sie*) als Leserkontakt. Ich orientiere mich dabei am POS-tag *PPER*. Reflexiv-Konstruktionen wie *Das knnt ihr euch beispielsweise an der Nordsee genauer anschauen* annotiere ich gemt Einschrnkung auf Personalpronomen der 2. Person nicht doppelt, sondern am Erstglied *ihr*.

Anreden, die innerhalb von Zitaten auftreten, wie zum Beispiel, *„Bist du irre? Das machst du freiwillig?“*, richten sich nicht an den Leser. Sie haben in dem Sinne keine auertextliche Funktion. Daher annotiere ich sie nicht als *anr*.

Erwartung

Aufgrund der Arbeiten aus den Bereichen der Sachtextforschung/Textlinguistik, Textverständlichkeitsmodelle und insbesondere der Registerforschung erwarthe ich, dass in der *GEOLino* Leserkontakt hergestellt wird. In der *GEO* hingegen dürfte entweder gar kein oder zumindest signifikant weniger Leserkontakt hergestellt werden als in der *GEOLino*. Treten hier tatsächlich signifikante Unterschiede auf, so ist das ein Indiz für einen Registerunterschied zwischen *GEOLino* und *GEO*.

4.2.13 Metasprachliche Elemente

Jeder Text weist eine Überschrift sowie eventuell einen kurzen Einleitungstext (Teaser) und eine oder mehrere Unterüberschriften auf. Dies sind vom Fließtext losgelöste metasprachliche Elemente und dienen dazu, den Text zu strukturieren. In dieser Funktion haben metasprachliche Elemente eine andere sprachliche Charakteristik als der Fließtext. So sind Überschriften meist fragmentarisch und spiegeln nicht die sprachliche Beschaffenheit des Fließtextes wider. Teaser stehen mehr oder weniger für sich und sollen den Leser mit wenigen und starken Mitteln dazu verleiten, weiter zu lesen. Auch sie unterscheiden sich sprachlich vom Fließtext, denn sie fassen Sachverhalte eher elliptisch zusammen und stehen in keinem chronologischen Zusammenhang zum folgenden Fließtext. Sie bilden eher eine eigene kurze Geschichte dessen, was gleich ausführlich erzählt wird und können darauf hinweisen, was den Leser im eigentlichen Artikel erwartet.

Aufgrund der sprachlichen Unterschiede zwischen Fließtext und Metasprachlichem muss ich metasprachliche Elemente separieren. Dazu annotiere ich Überschriften (Titel) als *t*, Unterüberschriften (Untertitel) als *ut* und Teaser als *tea* auf den Ebenen *titel*, *untertitel* und *teaser* und schließe diese Ebenen von weiterer Annotation und Analyse aus.

5. Analyse

In den folgenden Abschnitten stelle ich die Methode und die Ergebnisse meiner Analysen vor. Dazu formuliere ich zunächst allgemeine Erwartungen an den Vergleich der beiden Korpora. Ich gebe in Form einer Hauptkomponentenanalyse einen ersten optischen Eindruck des Korpusvergleichs. In einer Voranalyse zum univariaten Vergleich prüfe ich die Wirkung der Autoren- sowie der Geschlechts-Komponente, bevor ich die beiden Korpora hinsichtlich der einzelnen sprachlichen Merkmale gegenüberstelle. Eine Kritik soll abschließend auf problematische Aspekte der Analyse hinweisen.

5.1 Allgemeine Erwartungen

Anhand der Literatur und der daraus abgeleiteten sprachlichen Merkmale sowie der Richtlinien ergeben sich allgemeine Erwartungen an meine Arbeit. Zunächst erwarte ich, dass die sprachlichen Merkmale relevant für den Korpusvergleich sind, das heißt dass sich die Texte in der *GEOLino* und der *GEO* hinsichtlich dieser Merkmale unterscheiden. Die meisten Merkmale wurden von verschiedenen Autoren und aus unterschiedlichen Forschungsperspektiven betrachtet. Das lässt darauf schließen, dass diese Merkmale beim Lesen von Sachtexten eine Rolle spielen, sei es in Bezug auf das Verständnis, den kognitiven Arbeitsaufwand oder andere Aspekte. Insbesondere die Merkmale, die von den Autoren völlig entgegengesetzt eingeschätzt werden, wie zum Beispiel der Gebrauch von

Komparation, sind für meine Arbeit interessant: Ich erwarte, dass ich diese Kontroversen empirisch begründet auflösen kann, wenn auch nur in dem kleinen Rahmen des Vergleichs von *GEOlino* und *GEO*. Wenn die Merkmale relevant sind, erwarte ich, systematische sprachliche Unterschiede zwischen der *GEOlino* und der *GEO* zu finden. Weiterhin erwarte ich, dass diese Unterschiede statistisch signifikant sind.

In Bezug auf meine Fragstellung (2) – Liegen in *GEOlino* und *GEO* unterschiedliche Register vor? – ist sowohl das Merkmal *Leserkontakts* als auch das Verhältnis von attributiven zu prädikativen Adjektiven interessant. Aufgrund von Hypothese (1) erwarte ich auch hier signifikante Unterschiede zwischen *GEOlino* und *GEO*. Anhand dieser beiden Merkmale könnte man evtl. einen Ansatz für die Formulierung eines eigenen *GEOlino*-Registers wagen.

Insgesamt erwarte ich, dass ich durch meine Methode systematische sprachliche Unterschiede in der *GEOlino* und *GEO* aufdecken kann und die sprachliche Abgrenzung der *GEOlino* nach oben hin zur *GEO* schaffe.

5.2 Hauptkomponentenanalyse

Die Hauptkomponentenanalyse (HKA) zeigt eine Clusterung der Subkorpora. Mit einigen Überschneidungen clustern sie auf drei Arten. In Abbildung 1 stelle ich zusammen, welche Farben in den folgenden Abbildungen für welche Gruppen stehen. Dabei bezeichnen die Kürzel in der Mitte den jeweiligen Text, also *L6* für *GEOlino6*, *G7* für *GEO7* etc.









 L6	<i>GEOlino</i>	 G7	<i>GEO</i>
 G7	männliche Autoren	 L12	weibliche Autoren
 G5	männliche Autoren <i>GEO</i>	 G35	weibliche Autoren <i>GEO</i>
 L5	männliche Autoren <i>GEOlino</i>	 L6	weibliche Autoren <i>GEOlino</i>

Abbildung 1: Legende HKA

Abbildung 2 zeigt eine korpuspezifische Clusterung der Subkorpora. Bis auf die beiden Texte *GEOLino2* und *GEOLino16* lässt die HKA vermuten, dass es tatsächlich korpuspezifische sprachliche Unterschiede zwischen *GEOLino* und *GEO* gibt.

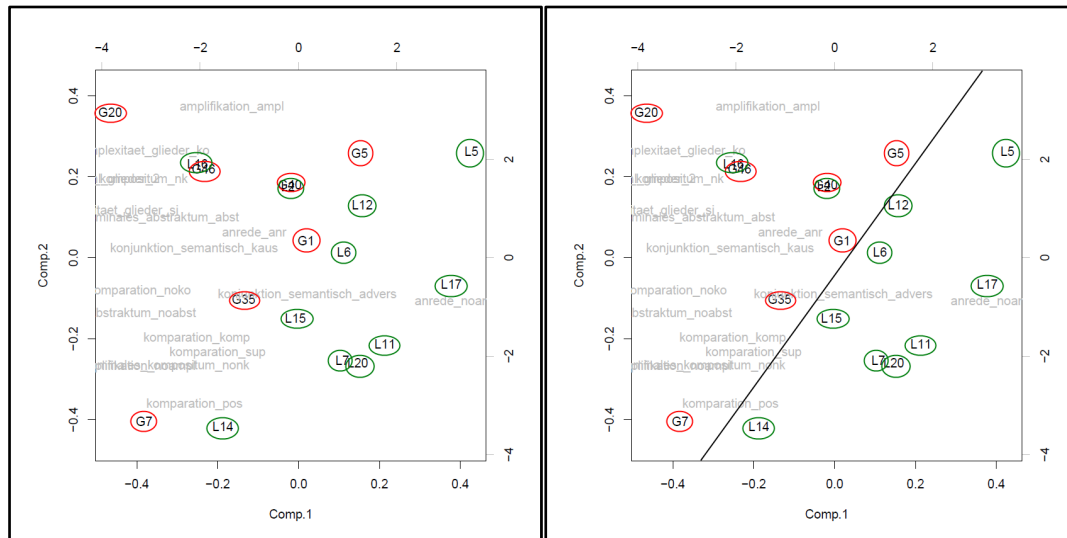


Abbildung 2: korpuspezifische Clusterung

In Abbildung 3 kann man erkennen, dass die Subkorpora auch geschlechtsspezifisch clustern und zwar Korpusübergreifend. Dies ist hier deutlicher für die Frauen zu erkennen als für die Männer. Die Clusterung ist zwar nicht so klar wie die korpuspezifische Clusterung. Sie ist jedoch ein Indiz dafür, dass die SV *Geschlecht* die AV beeinflusst und Unterschiede zwischen den Texten der beiden Korpora nicht eindeutig auf die UV *Korpus* zurückzuführen sind.

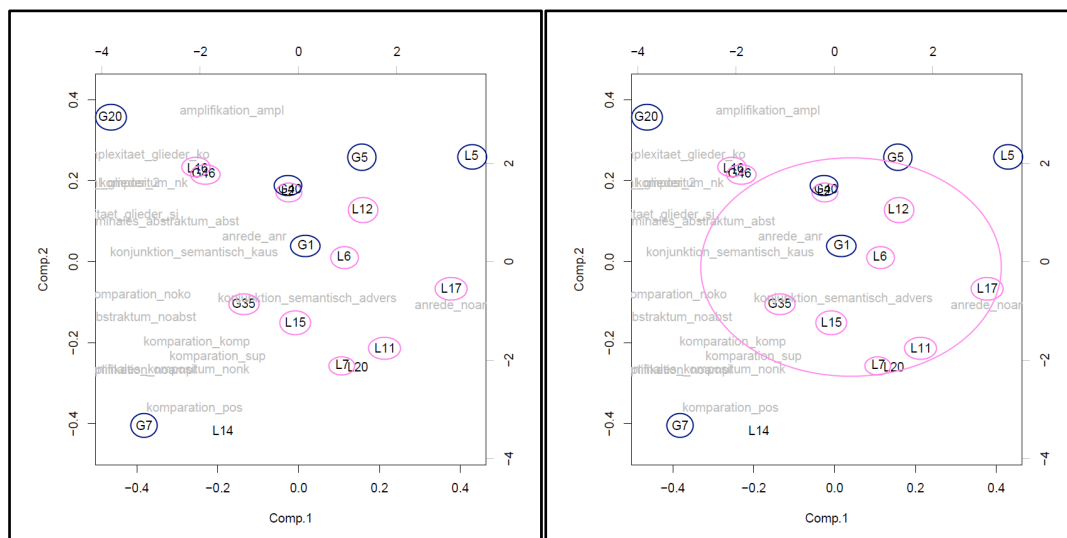


Abbildung 3: geschlechtsspezifische Clusterung

Abbildung 4 zeigt eine Kombination aus den beiden bisher betrachteten Clusterungen. Man erkennt, dass die Subkorpora sowohl korpus- als auch geschlechtsspezifisch clustern. Dabeizeigt sich die jeweilige sprachliche Reichweite der vier Geschlechtergruppen recht deutlich (der Text *L5* stellt dabei den einzigen Vertreter der Gruppe *GEOLino_m* dar). Innerhalb der Geschlechtergruppen *m* und *w* lässt sich nach wie vor der korpuspezifische Unterschied beobachten.

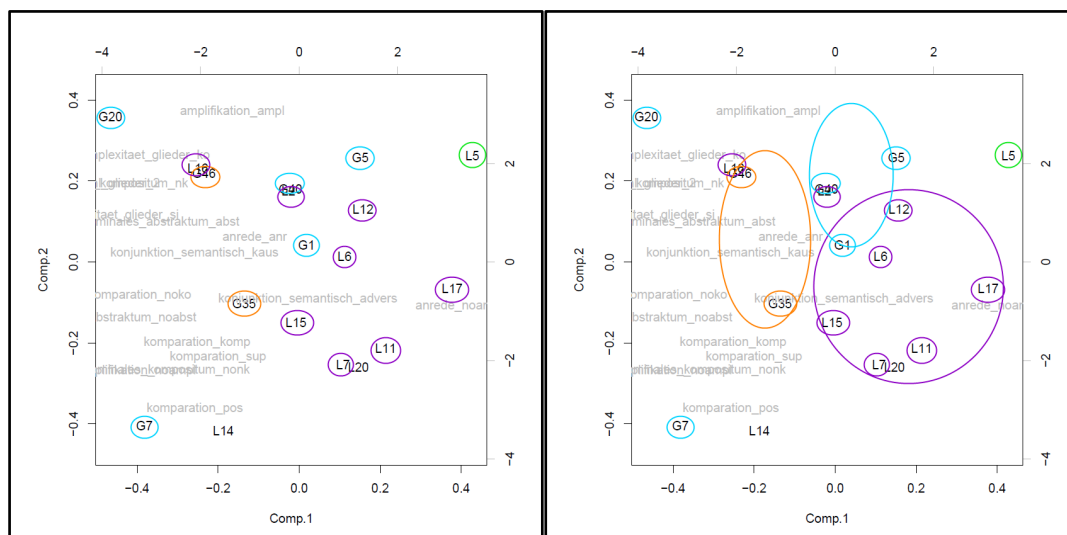


Abbildung 4: korpus- und geschlechtsspezifische Clusterung

Die HKA liefert einen ersten optischen Eindruck der Daten und zeigt, dass es relativ deutliche Unterschiede zwischen den *GEOLino*-Texten und den *GEO*-Texten gibt. Außerdem gibt sie Aufschluss darüber, dass sich die Texte auch geschlechtsspezifisch unterscheiden. Weil in beiden Fällen eine Clusterung zu erkennen ist, liegt die Vermutung nahe, dass es sich hier um unterschiedliche Register handeln könnte (vgl. Fragestellung (2)). Durch eine Faktoranalyse könnte man dieser Frage nun nachgehen. Aus technischen Gründen muss dies jedoch entfallen, weil die Daten hierfür noch eine zu hohe Redundanz für das entsprechende Programm aufweisen, das heißt, dass sich zu viele Merkmale entsprechen. So treten zum Beispiel *tags* auf den Annotationsebenen *anzahl.glieder* und *komplexitaet.glieder* ausnahmslos gemeinsam auf (Quelle: Julia Richling, persönliches Gespräch). Deswegen müssen an dieser Stelle die Abbildungen der HKA als erste Einordnung der Daten genügen.

5.3 Univariater Vergleich

Ich verwende für die Vor- und die Hauptanalyse das Programm *R*. Um eine bessere Lesbarkeit zu gewährleisten, habe ich alle Autorennamen durch Buchstaben ersetzt.

Im Zusammenhang mit der Korpuserstellung habe ich bereits darauf hingewiesen, dass die *GEOLino*-Texte sich hinsichtlich der Textlänge teilweise erheblich von den *GEO*-Texten unterscheiden. Unterschiedliche Textlänge kann zwar eine Quelle für Varianz sein, weil in längeren Texten potenziell mehr Vorkommen existieren. Trotzdem ist die Wahrscheinlichkeit hoch genug, dass die Ergebnisse sich nicht deutlich verändern würden, wenn dieser Unterschied nicht bestünde (Quelle: Felix Golcher, persönliches Gespräch).

In der Hauptanalyse prüfe ich für jedes einzelne Merkmal, ob sich seine Auftretenshäufigkeit (Frequenz) zwischen den Korpora signifikant unterscheidet. Dafür

vergleiche ich die Korpora mit dem *linear model* bzw. *summary(lm)*. Dadurch ist es möglich, mehrere UV gleichzeitig in die Analyse einzubeziehen – hier also *Korpus* und *Geschlecht*. In der Analyse sind diese Variablen mit *Corpus* und *Sex* bezeichnet. Durch das Modell kann ich prüfen, wie viel Einfluss eine UV auf die AV hat, wenn alle anderen UV bereits in die Funktion einbezogen wurden, ihre Wirkung also bereits in Betracht gezogen wurde. Ein einfacher T-Test ist für meine Analyse nicht geeignet, da dieser nur eine UV einbeziehen kann.

Weiterhin gibt das Modell Aufschluss über das Signal-Rausch-Verhältnis (SRV), das heißt über die Frage, wie viel der Varianz durch die Einbeziehung der Variablen *Korpus* und *Geschlecht* erklärt wird. Anders formuliert: Wie gut erklärt das Modell die beobachtete Varianz in den Daten?

5.3.1 Voranalyse

Eine Voranalyse soll zeigen, ob es eine SV *Autor* gibt, welche die Ergebnisse verfälschen kann (zur Begründung vgl. 4.1.3). Dazu führe ich einen korpusinternen Vergleich durch, indem ich innerhalb der Korpora jeweils die beiden Gruppen *m* und *w* aufstelle und statistisch prüfe, ob sich die Texte innerhalb dieser beiden Gruppen signifikant voneinander unterscheiden. Dabei bezeichnet jeweils *m* die Gruppe der Texte, die von männlichen Autoren verfasst wurden und *w* die Gruppe der Texte, die von weiblichen Autoren stammen. So entstehen die vier Gruppen *GEO_m*, *GEO_w*, *GEOLino_m* und *GEOLino_w*. Ich stelle hier die Frage:

- (1) Gibt es korpusinterne sprachliche Unterschiede innerhalb der Texte von männlichen bzw. weiblichen Autoren?

Sind keine signifikanten Unterschiede festzustellen, kann ich ausschließen, dass eine SV *Autor* die Ergebnisse der weiteren Analyse beeinflusst. In diesem Fall führe ich eine zweite Voranalyse durch, um den Einfluss der SV *Geschlecht* zu prüfen und stelle die Frage:

- (2) Gibt es korpusinterne sprachliche Unterschiede zwischen den Texten von männlichen und weiblichen Autoren?

Zunächst vergleiche ich gemäß Fragestellung (1) die Texte einer Geschlechtergruppe. Dazu führe ich einen Vergleich der Wortartenverteilung (POS-tags) durch.

Ich ermittle durch einen Chi²-Test, ob es in der Verteilung der POS-tags innerhalb der Texte der männlichen bzw. weiblichen Autoren in den beiden Korpora signifikante Unterschiede gibt. Dabei beschränke ich mich auf POS-tags, die mindestens 150 Mal über die jeweilige Gruppe summiert auftreten. Satzzeichen schließe ich dabei aus.

Das Ergebnis für die Gruppe *GEO_m* ist in Abbildung 5 dargestellt:

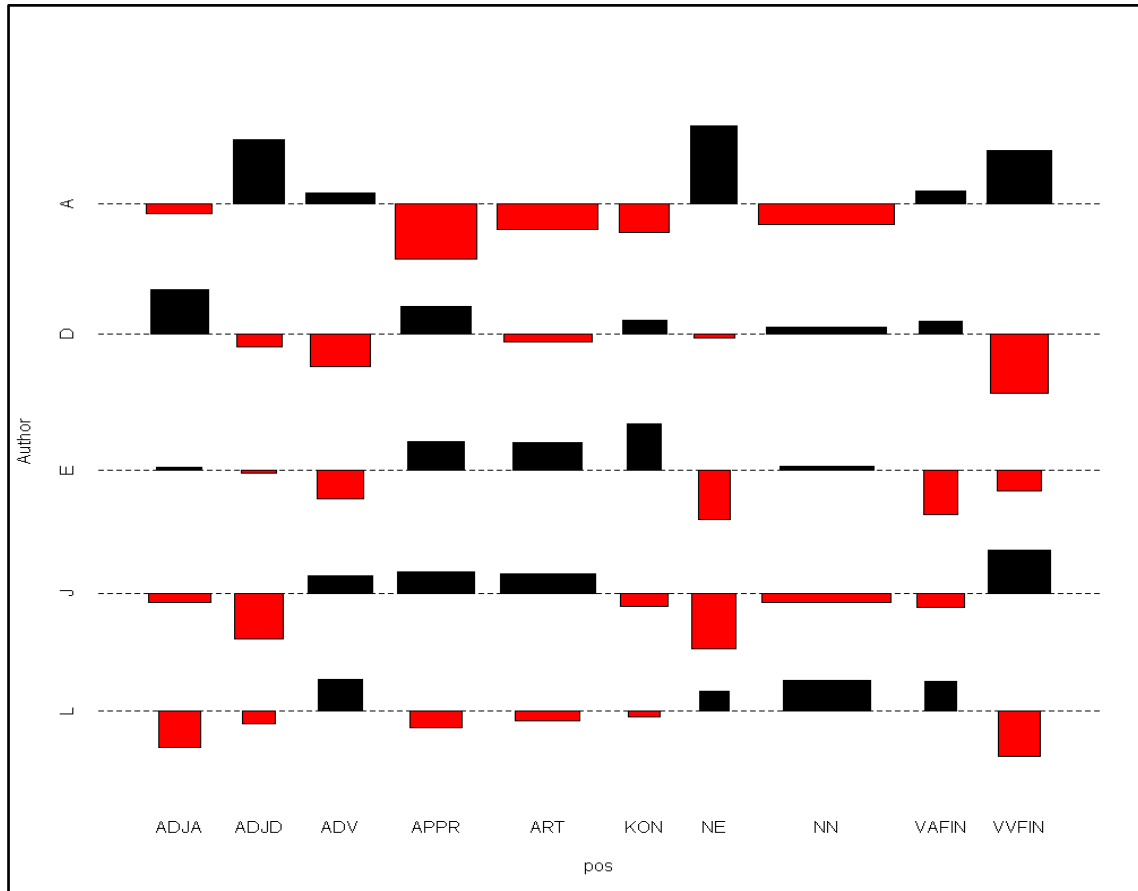


Abbildung 5: Verteilung der POS-tags in der Gruppe *GEO_m*

Jede Linie steht für den Text eines Autors (A, D, E, J und L). Horizontal darauf angeordnet sieht man die Ausprägungen der verschiedenen POS-tags. Dabei bezeichnen rote, nach unten gerichtete Balken eine niedrige Frequenz der POS-tags im Vergleich zu den anderen Texten. Je länger der Balken, desto seltener tritt der POS-tag im Text auf.

Schwarze, nach oben gerichtete Balken bezeichnen eine hohe Frequenz der POS-tags im Vergleich zu den anderen Texten. Je länger der Balken, desto häufiger tritt der POS-tag im Text auf. Die Breite der Balken spiegelt die Auftretenshäufigkeit des jeweiligen POS-tagswider.

Liest man die Balken vertikal (also spaltenweise), kann man den Gebrauch der einzelnen POS-tags in den Texten vergleichen. Dabei fällt auf, dass es keine vertikale Gleichverteilung gibt. Bei jedem POS-tag ergibt sich über die Texte hinweg eine Mischung aus höher- und niedrigerfrequenten Vorkommen. Das bedeutet, dass sich die Texte hier stark unterscheiden. Tatsächlich zeigt ein Chi²-Test, dass innerhalb der Gruppe *GEO_m* ein signifikanter Unterschied in der Verteilung der Wortarten vorliegt (Chi²= 67,1, df= 36, p= 0,001).

Ich kann meine Fragestellung (1) also positiv beantworten: Es gibt innerhalb der Korpora sprachliche Unterschiede innerhalb der Texte von männlichen bzw. weiblichen Autoren. Da ich bereits hier feststelle, dass es in mindestens einer der vier Gruppen – *GEO_m* – einen signifikanten Unterschied in der Wortartenverteilung gibt, erübrigt sich der Vergleich

innerhalb der anderen Gruppen. Der Vergleich ist für die Gruppe *GEOLino_m* ohnehin nicht möglich, da hier nur ein einziger Text vorhanden ist. Auch der Vergleich für die Gruppe *GEO_w* ist problematisch, da diese Gruppe nur zwei Texte enthält.

Weil sich hier gezeigt hat, dass es innerhalb der Gruppe *GEO_m* signifikante Unterschiede gibt, erübrigt sich auch der Vergleich zwischen *GEO_m* und *GEO_w*. Es ist nicht möglich, diesem Vergleich eine homogene Gesamtheit der männlichen Texte zugrunde zu legen. Daher unterscheiden sich die Texte von *GEO_m* und *GEO_w* zwangsläufig. Die zweite Voranalyse gemäß Fragestellung (2) kann daher entfallen.

Aus den Ergebnissen der Voranalyse folgt, dass ich die einzelnen Texte pro Korpus nicht als Grundgesamtheit zusammenfassend behandeln kann. Es gibt keinen „gemeinsamen Nenner“, also keinen echten Mittelwert. Die Texte liegen sprachlich einfach zu weit auseinander, unterscheiden sich zu stark voneinander. Aufgrund dieser großen Varianz muss ich jeden Text einzeln in die Analyse einbeziehen, jeder Text trägt also einen einzigen Datenpunkt zur entsprechenden Frage bei. Zudem beziehe ich die Variable *Geschlecht* als zweite UV neben der Variable *Korpus* in die Analyse ein, um für jedes Merkmal zu prüfen, ob es einen Geschlechtereffekt gibt, der für eventuelle Unterschiede verantwortlich ist.

5.3.2 Hauptanalyse

Ich möchte hier unbedingt betonen, dass die unterschiedliche Textanzahl in meinen Korpora problematisch für eine stichhaltige Analyse ist. Ebenso stellen das Ungleichgewicht in der Geschlechterverteilung und die Autoren-Komponente ein Problem dar. Daher muss ich meine Ergebnisse stark einschränken und darauf hinweisen, dass alle Rechnungen nur Näherungscharakter haben können. Durch den Einfluss der Autoren- und der Geschlechter-Komponente steigt zudem die Wahrscheinlichkeit, dass Unterschiede in den Daten zu gut durch den Zufall erklärt werden können, als dass man sie einem nicht-zufälligen Effekt zuschreiben könnte. Das bedeutet, dass trotz sichtbarer Unterschiede die Signifikanzwerte über der relevanten Grenze von 0,05 liegen können.

Zur besseren Übersicht formuliere ich in jedem Abschnitt noch einmal kurz die Erwartungen, die sich aus den Überlegungen für das jeweilige Merkmal ergeben. Entsprechende Grafiken füge ich im Anhang C bei.

5.3.2.1 Oberflächenmerkmale

In Bezug auf die Oberflächenmerkmale gehen aus der Literatur eindeutige Erwartungen hervor:

- (1) $\text{Wortlänge}_{\text{GEOLino}} < \text{Wortlänge}_{\text{GEO}}$
- (2) $\text{Satzlänge}_{\text{GEOLino}} < \text{Satzlänge}_{\text{GEO}}$

Ich ermittle die durchschnittliche Wortlänge für jeden Text in Buchstaben. Dabei zeigt sich, dass die durchschnittliche Wortlänge in jedem Text bei 8 bis 9 Zeichen liegt und zwar unabhängig vom Korpus und vom Geschlecht des Autors. Dabei variiert die Zeichenanzahl pro Wort bei *GEOLino* von 8,3 bis 8,8 und bei *GEO* von 8,2 bis 8,8. Die statistische Analyse zeigt, dass es hier keinen signifikanten Unterschied zwischen *GEOLino* und *GEO* gibt (t-Wert für die UV *Sex*=0,30, t-Wert für die UV *Corpus*=0,66). Die Wahrscheinlichkeit, dass die feinen Unterschiede zufällig sind, ist zu hoch, um Erwartung

(1) eindeutig bestätigen zu können. Das SRV besagt, dass lediglich 8% der Varianz durch das Modell erklärt werden.

Ich betrachte die durchschnittliche Satzlänge pro Text in der Anzahl der verwendeten Wörter zwischen zwei Satzendzeichen. Es stellt sich heraus, dass die durchschnittliche Satzlänge in der *GEOlino* mit ungefähr 10 bis fast 20 Wörtern sehr variiert. In der *GEO* liegt die Varianz bei 14 bis fast 20 Wörtern pro Satz. Dabei variieren die *GEOlino*-Texte weiter nach unten hin, während die *GEO*-Texte eher längere Sätze aufweisen. Dies würde Erwartung (2) bestätigen. Die statistische Analyse zeigt, dass es tatsächlich signifikante Unterschiede zwischen *GEOlino* und *GEO* im Hinblick auf die Satzlänge gibt. Diese Unterschiede treten sowohl korpus- als auch geschlechtsspezifisch auf (t-Wert für die UV *Sex*=0,04, t-Wert für die UV *Corpus*=0,01). Erwartung (2) wird somit von den Daten bestätigt. Aus dem SRV geht hervor, dass das Modell 39% der Varianz erklärt.

5.3.2.2 Wortbildung: Komposition

Aus der Literatur geht folgende Erwartung hervor:

$$(3) \text{nk}_{\text{GEOlino}} < \text{nk}_{\text{GEO}}$$

Ich ermittle das Verhältnis der nominalen Komposita (*nk*) zu den nicht zusammengesetzten und den lexikalisierten Substantiven (*nonk*). Dabei zeigt sich, dass das Verhältnis von *nk* zu *nonk* variiert – von ungefähr 0,1 bis 0,4 Vorkommen von *nk* pro *nonk* in den *GEOlino*-Texten bzw. 0,2 bis 0,5 in den *GEO*-Texten. Die statistische Analyse ergibt jedoch keinen signifikanten Unterschied zwischen *GEOlino* und *GEO* (t-Wert für die UV *Sex*=0,76, t-Wert für die UV *Corpus*=0,20). Die beobachteten Unterschiede sprechen für Erwartung (3). Die Wahrscheinlichkeit, dass die Unterschiede in der Verwendung von nominalen Komposita zufällig sind, ist jedoch zu hoch, um Erwartung (3) eindeutig bestätigen zu können. Laut SRV werden 22% der Varianz durch das Modell erklärt.

5.3.2.3 Anzahl der Hauptglieder

Die Angaben aus der Literatur lassen Erwartung (4) zu. Die Beschränkung auf Kompositions-Hauptglieder legen jedoch auch Erwartung (5) nahe, die hier sicher die stärkere ist:

$$(4) \text{Anzahl Glieder}_{\text{GEOlino}} < \text{Anzahl Glieder}_{\text{GEO}}$$

$$(5) \text{Anzahl Glieder}_{\text{GEOlino}} = \text{Anzahl Glieder}_{\text{GEO}}$$

Bereits ein erster Blick in die Daten zeigt, dass eine statistische Analyse hier nicht nötig ist: Alle nominalen Komposita in allen Texten bestehen aus zwei Hauptgliedern. Somit wird Erwartung (5) bestätigt, Erwartung (4) jedoch nicht.

5.3.2.4 Komplexität der Hauptglieder

Aus der Literatur sowie aus der Weiterführung einiger Annahmen ergibt sich die folgende Erwartung:

$$(6) \text{Komplexität Hauptglieder}_{\text{GEOlino}} < \text{Komplexität Hauptglieder}_{\text{GEO}}$$

Ich ermittle das Verhältnis von komplexen (*ko*) zu simplexen (*si*) Hauptgliedern. Dabei zeigt sich, dass das Verhältnis in den *GEOlino*-Texten mit ungefähr 0,2 bis 1,8

Vorkommen von *ko pro si* und in den *GEO*-Texten mit fast 0,7 bis 1,5 Vorkommen variiert. Dabei fallen die meisten *GEolino*-Texte in den Bereich unter 0,7 und damit unter den Mindestwert der *GEO*-Texte. Das bedeutet, dass in den *GEolino*-Texten eher weniger komplexe Hauptglieder verwendet werden und dies spricht für Erwartung (6). Aber auch hier ergibt die statistische Analyse keinen signifikanten Unterschied zwischen *GEolino* und *GEO* (t-Wert für die UV *Sex*=0,07, t-Wert für die UV *Corpus*=0,49). Doch zeigt sich hier tendenziell ein geschlechtsspezifischer Unterschied. Die Wahrscheinlichkeit, dass die Unterschiede in der Verwendung von komplexen Hauptgliedern zufällig sind, ist zu hoch, um Erwartung (6) eindeutig bestätigen zu können. Laut SRV werden 24% der Varianz durch das Modell erklärt.

5.3.2.5 Adjektivfrequenz

Aus der Literatur ergeben sich entgegengesetzte Erwartungen für den Gebrauch von Adjektiven. Je nachdem, welchem Ansatz man folgt, lassen sich die folgenden Erwartungen formulieren:

(7) Anzahl Adjektive_{GEolino} > Anzahl Adjektive_{GEO}

(8) Anzahl Adjektive_{GEolino} < Anzahl Adjektive_{GEO}

Ich ermittle den Anteil der Adjektive (*ADJA*, *ADJD*) an der Gesamt-Tokenzahl des Textes. Dabei fällt auf, dass sich *GEolino14* mit einem Anteil von 0,10 Vorkommen von *ADJA* bzw. *ADJD* pro Token nach oben hin Korpus übergreifend von allen anderen Texten abhebt. Hier ist die Adjektiv-Frequenz also vergleichsweise hoch. Insgesamt streuen die restlichen *GEolino*-Texte mit einem Anteil von ca. 0,06 bis ca. 0,07 weniger als die *GEO*-Texte. Diese weisen einen *ADJA*- bzw. *ADJD*-Anteil von ca. 0,06 bis 0,09 auf. Sieht man von *GEolino14* ab, würden diese Unterschiede für Erwartung (8) und gegen Erwartung (7) sprechen. Doch auch hier zeigt die statistische Analyse keinen signifikanten Unterschied (t-Wert für die UV *Sex*=0,58, t-Wert für die UV *Corpus*=0,42). Die Wahrscheinlichkeit, dass die Unterschiede in der allgemeinen Adjektivfrequenz zufällig sind, ist zu hoch, um Erwartung (7) oder (8) eindeutig bestätigen zu können. Laut SRV werden 16% der Varianz durch das Modell erklärt.

5.3.2.6 Attributive und prädikative Adjektive

Aus der Literatur lässt sich folgende Erwartung formulieren:

(9) Anzahl attributiver Adjektive_{GEolino} < Anzahl attributiver Adjektive_{GEO}

Ich ermittle das Verhältnis von attributiven Adjektiven (*ADJA*) zu prädikativen Adjektiven (*ADJD*).¹⁸ Dabei zeigt sich, dass das Verhältnis im Korpus *GEolino* mit ungefähr 1,0 bis 3,0 Vorkommen von *ADJA* pro *ADJD* mehr streut als im Korpus *GEO* (1,2 bis 2,6). Auch eine geschlechtsspezifische Gruppierung der Punkte lässt sich – zumindest korpusintern – erkennen: Die männlichen Autoren verwenden jeweils weniger attributive Adjektive als die weiblichen Autoren (wobei zu beachten ist, dass der männliche Anteil an den *GEolino*-Texten lediglich durch einen einzigen Autor vertreten wird). Doch auch hier zeigt die statistische Analyse keinen signifikanten Unterschied (t-Wert für die UV *Sex*=0,05, t-Wert für die UV *Corpus*=0,17). Die Wahrscheinlichkeit, dass die hier relativ feinen

¹⁸ Der pos-tag *ADJD* wird sowohl für prädikative als auch für adverbiale Adjektive verwendet.

Unterschiede in der Verwendung von attributiven Adjektiven zufällig sind, ist zu hoch, um Erwartung (9) eindeutig bestätigen zu können. Wie schon beim Merkmal *Komplexität Glieder* zeigt sich auch hier ein tendenzieller geschlechtsspezifischer Unterschied, der mit $t=0,05$ sogar fast signifikant ist. Laut SRV werden 25% der Varianz durch das Modell erklärt.

5.3.2.7 Komparation

In Bezug auf den Gebrauch von Komparation lassen sich aus der Literatur gegensätzliche Erwartungen ableiten:

$$(10) \text{Anzahl Superlative}_{GEOLino} < \text{Anzahl Superlative}_{GEO}$$

$$(11) \text{Anzahl Komparationen}_{GEOLino} > \text{Anzahl Komparationen}_{GEO}$$

Für Erwartung (10) setze ich die Vorkommen vom Superlativ (*sup*) ins Verhältnis zu den beiden anderen Komparationsstufen Positiv (*pos*) und Komparativ (*komp*). Dabei zeigt sich, dass die *GEOLino*-Texte mit 0,0 bis 0,25 Vorkommen von *sup* pro *pos* bzw. *komp* nach oben hin weiter streuen als die *GEO*-Texte (0,0 bis 0,08 Vorkommen). Damit werden in der *GEOLino* potenziell mehr Superlative verwendet als in der *GEO*. Dies spricht gegen Erwartung (10). Dazu muss jedoch gesagt werden, dass sich *GEOLino2* und *GEOLino14* mit einem Superlativ-Anteil von jeweils 0,25 Vorkommen nach oben hin Korpus übergreifend von allen anderen Texten abheben. Die statistische Analyse ergibt keinen signifikanten Unterschied (t-Wert für die UV *Sex*=0,14, t-Wert für die UV *Corpus*=0,78). Die Wahrscheinlichkeit, dass die Unterschiede in der Verwendung des Superlativs gegenüber den anderen beiden Komparationsstufen zufällig sind, ist zu hoch, um Erwartung (10) eindeutig bestätigen oder widerlegen zu können. Laut SRV werden 27% der Varianz durch das Modell erklärt.

Für Erwartung (11) setze ich die beiden Komparationsstufen *komp* und *sup* ins Verhältnis mit *pos*. Dabei zeigt sich, dass die Werte in der *GEOLino* mit 0,14 bis 0,35 Vorkommen von *komp* bzw. *sup* pro *pos* weniger streuen als in der *GEO* (0,06 bis 0,36). Insgesamt werden in der *GEOLino* eher mehr Komparative und Superlative verwendet als in der *GEO*. Dies spricht für Erwartung (11). Aber auch hier ergibt die statistische Analyse keinen signifikanten Unterschied (t-Wert für die UV *Sex*=0,54, t-Wert für die UV *Corpus*=0,92). Die Wahrscheinlichkeit, dass die Unterschiede in der Verwendung von Komparativen und Superlativen zufällig sind, ist zu hoch, um Erwartung (11) eindeutig bestätigen zu können. Laut SRV werden lediglich 3% der Varianz durch das Modell erklärt.

5.3.2.8 Adjektivische Amplifikation

Für den Gebrauch von adjektivischen Amplifikationen lassen sich zwei entgegengesetzte Erwartungen formulieren:

$$(12) \text{Anzahl adjektivischer Amplifikation}_{GEOLino} < \text{Anzahl adjektivischer Amplifikation}_{GEO}$$

$$(13) \text{Anzahl adjektivischer Amplifikation}_{GEOLino} > \text{Anzahl adjektivischer Amplifikation}_{GEO}$$

Ich setze die Anzahl von adjektivischen Amplifikationen (*ampl*) ins Verhältnis zur Anzahl nicht-amplifizierter Adjektive (*noampl*). Es fällt auf, dass in fast einem Drittel der

GEOLino-Texte überhaupt keine Amplifikation verwendet wurde, in den *GEO*-Texten hingegen finden sich keine absoluten Nullwerte. Die Werte in den *GEOLino*-Texten variieren vom absoluten Nullwert 0,0 bis 0,04 Vorkommen von *ampl* pro *noampl*, in den *GEO*-Texten von 0,005 bis 0,03. Die Nullwerte in der *GEOLino* sprechen für Erwartung (12). Die statistische Analyse zeigt jedoch, dass es keinen signifikanten Unterschied gibt (t-Wert für die UV *Sex*=0,56, t-Wert für die UV *Corpus*=0,73). Die Wahrscheinlichkeit, dass die Unterschiede in der Verwendung adjektivischer Amplifikationen zufällig sind, ist zu hoch, um Erwartung (12) oder (13) eindeutig bestätigen zu können. Laut SRV werden lediglich 2% der Varianz durch das Modell erklärt.

5.3.2.9 Nominale Abstrakta

Aus der Literatur geht eindeutig die Erwartung hervor:

$$(14) \text{ Anzahl nominale Abstrakta}_{GEOLino} < \text{Anzahl nominale Abstrakta}_{GEO}$$

Ich setze – ähnlich wie Groeben (1982:226) – die Anzahl der nominalen Abstrakta (*abst*) ins Verhältnis zu den restlichen Substantiven (*noabst*). Hier zeigt sich, dass die *GEOLino*-Texte mit 0,03 bis 0,15 Vorkommen von *abst* pro *noabst* weniger Abstrakta aufweisen als die *GEO*-Texte mit 0,10 bis 0,16 Vorkommen. Dies spricht für Erwartung (14). Nach Groebens (ebd.) Liste wären die *GEOLino*-Texte mit 3 bis 15% Abstrakta sehr bis mittelmäßig konkret, die *GEO*-Texte mit 10 bis 16% konkret bis mittelmäßig konkret. Auch wenn eine Tendenz zu erkennen ist, ergibt die statistische Analyse keinen signifikanten Unterschied (t-Wert für die UV *Sex*=0,99, t-Wert für die UV *Corpus*=0,18). Die Wahrscheinlichkeit, dass die Unterschiede in der Verwendung von nominalen Abstrakta zufällig sind, ist zu hoch, um Erwartung (14) eindeutig bestätigen zu können. Laut SRV werden 19% der Varianz durch das Modell erklärt.

5.3.2.10 Konjunktionsfrequenz

Aus der Literatur lassen sich gegensätzliche Erwartungen für den Gebrauch von Konjunktionen formulieren:

$$(15) \text{ Anzahl Konjunktionen}_{GEOLino} > \text{Anzahl Konjunktionen}_{GEO}$$

$$(16) \text{ Anzahl Konjunktionen}_{GEOLino} < \text{Anzahl Konjunktionen}_{GEO}$$

Um diese Erwartungen zu prüfen, setze ich die Anzahl der Konjunktionen (*KOUI*, *KOUS*, *KON*, *KOKOM*) ins Verhältnis zur Gesamt-Tokenzahl des jeweiligen Textes. Hier zeigt sich, dass die Werte bei den *GEOLino*-Texten mit 0,03 bis 0,06 Vorkommen von *KOUI*, *KOUS*, *KON* bzw. *KOKOM* pro Token weiter streuen als bei den *GEO*-Texten mit 0,03 bis 0,05 Vorkommen. Dabei scheinen die Werte tendenziell höher angesetzt als die der *GEO*. Das würde für eine leicht höherfrequente Verwendung von Konjunktionen in der *GEOLino* und damit für Erwartung (15) sprechen. Doch auch hier ergibt die statistische Analyse keinen signifikanten Unterschied (t-Wert für die UV *Sex*=0,92, t-Wert für die UV *Corpus*=0,41). Die Wahrscheinlichkeit, dass die Unterschiede in der allgemeinen Verwendung von Konjunktionen zufällig sind, ist zu hoch, um Erwartung (15) oder (16) eindeutig bestätigen zu können. Laut SRV werden lediglich 7% der Varianz durch das Modell erklärt.

5.3.2.11 Semantische Klassen von Konjunktionen

In Bezug auf den Gebrauch von adversativen und kausalen Konjunktionen gehen unterschiedliche Erwartungen aus der Literatur hervor:

- (17) Anzahl adversativer und kausaler Konjunktionen_{GEOLino} > Anzahl adversativer und kausaler Konjunktionen_{GEO}
- (18) Anzahl adversativer und kausaler Konjunktionen_{GEOLino} = Anzahl adversativer und kausaler Konjunktionen_{GEO}

Ich habe sowohl adversative als auch kausale Konjunktionen anhand einer Liste annotiert und mich damit nicht nach bestimmten POS-*tags* gerichtet. Sie tragen nicht konsequent die gleichen POS-*tags*, daher macht es keinen Sinn, sie ins Verhältnis zu den POS-*tags* der Konjunktionen zu setzen. Deswegen setze ich hier jeweils die Vorkommen adversativer und kausaler Konjunktionen (*advers*, *kaus*) ins Verhältnis zur Gesamt-Tokenzahl.

Beim Vergleich der Vorkommen von *advers* fällt auf, dass die *GEOLino*-Texte mit 0,001 bis 0,013 Vorkommen pro Token weiter streuen als die *GEO*-Texte mit 0,002 bis 0,005 Vorkommen. Außerdem gibt es mit *GEOLino20* einen Text, der sich nach oben hin Korpus übergreifend von allen anderen Texten absetzt. Selbst wenn man von *GEOLino20* absieht, werden adversative Konjunktionen in der *GEOLino* öfter verwendet als in der *GEO*. Dies spricht – in Bezug auf adversative Konjunktionen – für Erwartung (17). Die statistische Analyse zeigt jedoch, dass es keinen signifikanten Unterschied gibt (t-Wert für die UV *Sex*=0,10, t-Wert für die UV *Corpus*=0,98). Die Wahrscheinlichkeit, dass die Unterschiede in der Verwendung von adversativen Konjunktionen zufällig sind, ist zu hoch, um Erwartung (17) eindeutig bestätigen zu können. Erwartung (18) wird allein durch die Signifikanzwerte, nicht jedoch durch die Varianzbeobachtung bestätigt. Laut SRV werden 27% der Varianz durch das Modell erklärt.

Der Vergleich von *kaus* zeigt für beide Korpora nach oben hin ganz ähnliche Spitzenwerte (ungefähr 0,004 Vorkommen von *kaus* pro Token). Nach unten hin unterscheiden sich die Korpora etwas deutlicher mit 0,000 Vorkommen in *GEOLino17* und 0,001 Vorkommen in *GEO5*. Doch auch hier lassen sich keine signifikanten Unterschiede feststellen (t-Wert für die UV *Sex*=0,51, t-Wert für die UV *Corpus*=0,48). Was für die adversativen Konjunktionen gilt, gilt hier entsprechend: Die Wahrscheinlichkeit, dass die feinen Unterschiede in der Verwendung von kausalen Konjunktionen zufällig sind, ist zu hoch, um Erwartung (17) eindeutig bestätigen zu können. Erwartung (18) wird durch die Signifikanzwerte hingegen bestätigt. Laut SRV werden 4% der Varianz durch das Modell erklärt.

5.3.2.12 Leserkontakt

Aus der Literatur zur Herstellung von Leserkontakt bzw. zum Gebrauch von direkter Anrede an den Leser geht folgende Erwartung hervor:

- (19) Vorkommen direkter Anrede_{GEOLino} > Vorkommen direkter Anrede_{GEO}

Ich setze die Vorkommen von direkter Anrede an den Leser durch Personalpronomen der 2. Person (*anr*) ins Verhältnis zu den Vorkommen der restlichen Personalpronomen (*noanr*). Hier fällt auf, dass mit vier Texten fast ein Drittel der *GEOLino*-Texte gar keine direkte Anrede an den Leser aufweist, bei den *GEO*-Texten sind es sogar sechs von sieben. *GEOLino16* hebt sich mit 0,8 Vorkommen von direkter Anrede pro *noanr* nach oben hin

von allen anderen Texten ab. In den restlichen Texten variieren die Vorkommen von 0,00 bis 0,13 (*GEOLino*) bzw. bis 0,03 (*GEO*). In der *GEOLino* wird die direkte Anrede also öfter verwendet als in der *GEO*. Das spricht für Erwartung (19). Die statistische Analyse zeigt jedoch keinen signifikanten Unterschied (t-Wert für die UV *Sex*=0,67, t-Wert für die UV *Corpus*=0,58). Die Wahrscheinlichkeit, dass die Unterschiede in der Verwendung von direkter Anrede an den Leser zufällig sind, ist zu hoch, um Erwartung (19) eindeutig bestätigen zu können, auch wenn die Varianzbeobachtung darauf hinweist. Laut SRV werden lediglich 9% der Varianz durch das Modell erklärt.

GEOLino16 hebt sich zwar nicht signifikant von allen anderen Texten in Bezug auf die Frequenz der direkten Anrede ab. Doch ich möchte hier darauf hinweisen, dass in *GEOLino16* ein potenziell anderes Register verwendet werden könnte als in den anderen Texten: Der Text ist eine Mischung aus informierendem Sachtext und fiktivem Reiseprospekt. Durch diese zusätzliche Komponente könnte sich die erhöhte Frequenz von direkter Anrede ergeben, denn in Reiseprospekten oder -führern wird der Leser vielleicht eher direkt angesprochen als in reinen informativen Sachtexten.

5.3.3 Kritik

Wie bereits erwähnt müssen die Ergebnisse in einem sehr engen Rahmen gesehen und als Näherungswerte verstanden werden. Ich habe versucht, möglichst gleichartige Korpora zusammenzustellen. Trotzdem ist es mir nicht gelungen, Korpora zu generieren, deren Texte als Gesamtheit zusammengefasst werden können und als Vergleichsgrundlage dienen. Die Autor- sowie die Geschlechter-Komponente haben einen so starken Einfluss auf die AV, dass stets eine zu hohe Zufalls-Wahrscheinlichkeit angenommen werden muss, um die beobachteten Unterschiede zu erklären. Somit stellen sich zwei Fragen: Wie aussagekräftig ist die Analyse und wie geeignet sind meine Daten noch für die ursprüngliche Forschungsfrage?

Man kann auch bezweifeln, dass die Merkmale, die in der Literatur als relevant angesehen werden, dies tatsächlich sind. Viele Autoren haben nicht wissenschaftlich bzw. introspektiv argumentiert, daher gibt es durchaus eine gewisse Wahrscheinlichkeit, dass die Merkmale keine oder keine große Rolle in der Unterscheidung zwischen *GEOLino* und *GEO* spielen. Weiterhin kann es sein, dass sich die *GEOLino* nicht an diese sprachlichen „Vorgaben“ hält. Ich bin in meiner Arbeit davon ausgegangen, dass die *GEOLino* ihren Lesern sprachlich „entgegenkommt“, das heißt sprachliche Mittel einsetzt, die das Verständnis erleichtern bzw. das Interesse des Lesers am Text wecken und aufrechterhalten. Auch dies muss nicht unbedingt eingehalten werden und kann daher zu den berichteten Ergebnissen führen.

6. Zusammenfassung

Die HKA hat gezeigt, dass es durchaus systematische sprachliche Unterschiede zwischen *GEOLino* und *GEO* gibt. Dieses Ergebnis habe ich durch den univariaten Vergleich nicht replizieren können. Den einzigen signifikanten Unterschied zwischen *GEOLino* und *GEO* habe ich für die Satzlänge feststellen können: Demnach werden in der *GEOLino* kürzere Sätze verwendet als in der *GEO*. Für alle anderen untersuchten Merkmale ergab die statistische Analyse keine signifikanten Unterschiede. Ich konnte Varianzbeobachtungen zu den Bereichen *Komplexität der Hauptglieder*, *Adjektivfrequenz*,

Komparation, Adjektivische Amplifikation, Nominale Abstrakta, Konjunktionsfrequenz, adversative Konjunktionen sowie *Leserkontakt* anstellen und tendenzielle Unterschiede benennen.

Ich habe Merkmale untersucht, die in der Literatur als relevant für das erfolgreiche Lesen von Sachtexten angesehen werden. Entgegen dieser Annahme scheinen die meisten dieser Merkmale keine große distinktive Rolle in *GEolino* und *GEO* zu spielen.

Für die Wort- und Satzlänge geht aus der Literatur eine Präferenz für kurze Formulierungen hervor, insbesondere von Kommerell (2008), Christmann/Groeben (2006), Baurmann (2009) und Groeben (1982). Demnach erwartete ich in den *GEolino*-Texten kürzere Wörter und Sätze als in den *GEO*-Texten. Dies habe ich statistisch jedoch nur für die Satzlänge nachweisen können. In Bezug auf die Wortlänge haben sich keine Unterschiede gezeigt. Doch lassen sich hier Varianzen beobachten, die andeuten, dass in der *GEolino* tendenziell kürzere Wörter verwendet werden als in der *GEO*.

Im Bereich der Komposition betonte Gierlich (2005) die Verständnishemmende Wirkung von Komposita in Sachtexten. Davon und von den Ausführungen zur Wortlänge ausgehend sowie in der Annahme, dass Texte für junge Leser mehr Verständnisförderndes sprachliches Material aufweisen als Texte für Erwachsene, habe ich einen Unterschied in der Frequenz von Komposita erwartet: *GEolino*-Texte würden demnach weniger Komposita enthalten als *GEO*-Texte. Dies konnte ich nicht bestätigen.

Durch die Beschränkung auf Kompositions-Hauptglieder habe ich erwartet, dass sich die Anzahl der verwendeten Glieder pro Kompositum zwischen *GEolino* und *GEO* nicht unterscheidet. Dies konnte ich bestätigen.

In Bezug auf die Komplexität der Hauptglieder habe ich in Anlehnung an Kommerell (2008) sowie an die Ausführungen zur Wortlänge erwartet, dass in der *GEolino* weniger komplexe Hauptglieder auftreten als in der *GEO*. Auch dies habe ich nicht bestätigen können. Ich habe jedoch eine Varianz beobachten können, die nahelegt, dass in der *GEolino* tendenziell weniger komplexe Hauptglieder verwendet werden als in der *GEO*. Weiterhin habe ich eine Tendenz zum geschlechtsspezifischen Unterschied in der Verwendung von komplexen Hauptgliedern feststellen können.

Für den Gebrauch von Adjektiven gibt es in der Literatur gegensätzliche Annahmen von May (1996) und Kommerell (2008). Ich habe erwartet, diesen Gegensatz im kleinen Rahmen von *GEolino* und *GEO* empirisch entscheiden zu können. Diese Erwartung hat sich nicht erfüllt: Ich habe keine signifikanten Unterschiede in der Gebrauchsfrequenz von Adjektiven feststellen können. Doch auch hier variieren die Werte so, dass man einen tendenziellen Unterschied erkennen kann: Tendenziell werden in der *GEolino* weniger Adjektive verwendet als in der *GEO*.

Der Gebrauch von attributiven Adjektiven bzw. das Verhältnis von attributiven zu prädikativen Adjektiven spielt sowohl für meine Fragestellung (1) als auch (2) (siehe Abschnitt 1.) eine Rolle. Hier habe ich aufgrund der Annahmen aus der Literatur (Doderer (1960), Biber (1995)) eine geringere Frequenz in der *GEolino* als in der *GEO* erwartet. Auch diese Erwartung konnte ich statistisch nicht bestätigen. Ich konnte jedoch einen geschlechtsspezifischen Unterschied feststellen, der mit $t=0,05$ fast signifikant ist.

Im Bereich der Komparation sowie für den Gebrauch von adjektivischen Amplifikationen gibt es in der Literatur gegensätzliche Annahmen, unter anderem von Kommerell (2008) und von Gierlich (2005). Wie auch im Zusammenhang mit dem allgemeinen

Adjektivgebrauch habe ich erwartet, den Gegensatz empirisch im Rahmen von *GEOLino* und *GEO* auflösen zu können. Doch weder für den Gebrauch des Superlativs noch für den Gebrauch von Komparation allgemein sowie für den Gebrauch adjektivischer Amplifikationen konnte ich signifikante Unterschiede feststellen. Für den Gebrauch des Superlativs bzw. des Komparativs und Superlativs habe ich jedoch jeweils eine tendenziell höhere Frequenz in der *GEOLino* als in der *GEO* erkennen können. Außerdem zeigt sich in den Daten eine Varianz, die darauf schließen lässt, dass in der *GEOLino* tendenziell weniger adjektivische Amplifikationen verwendet werden als in der *GEO*.

Aus der Literatur geht eindeutig hervor, dass Abstrakta für junge Leser zu vermeiden sind, da sie kognitiv zu komplex sind und eher Verständnishemmend wirken (vgl. zum Beispiel DuBay (2004), Baurmann (2009) oder Kommerell (2008)). Daher habe ich in der *GEOLino* eine niedrigere Frequenz von Abstrakta erwartet als in der *GEO*. Doch auch hier habe ich keine signifikanten Unterschiede feststellen können. Tendenziell lassen sich jedoch auch hier Unterschiede beobachten: Die *GEOLino*-Texte weisen eher weniger Abstrakta auf als die *GEO*-Texte.

Im Bereich der Konjunktionen gibt es von Willenberg (2005) und Groeben (1982) gegensätzliche Annahmen in der Literatur. Auch hier habe ich erwartet, den Gegensatz im Rahmen von *GEOLino* und *GEO* empirisch aufzulösen. Doch auch hier konnte ich keine signifikanten Unterschiede in der Konjunktionsfrequenz feststellen. Betrachtet man die Varianzen, erkennt man jedoch, dass in der *GEOLino* tendenziell mehr Konjunktionen verwendet werden als in der *GEO*.

Bezüglich adversativer sowie kausaler Konjunktionen gibt es ebenfalls gegensätzliche Annahmen in der Literatur (vgl. Hiller (2010), Feilke (1996)). Diesen Gegensatz konnte ich zwar im Rahmen des Vergleichs von *GEOLino* und *GEO* statistisch für Feilke entscheiden, doch lassen sich im Bereich der adversativen Konjunktionen Varianzen beobachten, die wiederum gegen Feilkes Annahmen sprechen und vielmehr Hiller stützen: In der *GEOLino* werden tendenziell mehr adversative Konjunktionen verwendet als in der *GEO*.

Aus der Literatur geht eindeutig hervor, dass die direkte Anrede an den Leser gerade für junge Leser förderlich ist (vgl. zum Beispiel Baurmann (2009), Schulz/Tausch/Langer (2011) und Kommerell (2008)) bzw. Register unterscheidende Funktion hat (vgl. zum Beispiel Biber/Conrad/Reppen (2002)). Aufgrund dieser Annahme sowie der bereits erwähnten Annahme, Texte für junge Leser enthalten mehr sprachliches Material, welches das Verständnis fördert, habe ich in den *GEOLino*-Texten eine höhere Frequenz von direkter Anrede erwartet. Ich habe auch hier keine signifikanten Unterschiede feststellen können. Doch konnte ich hier Varianzen beobachten, die schon dafür sprechen, dass in der *GEOLino* tendenziell mehr direkte Anrede verwendet wird als in der *GEO*.

In der Gesamtbetrachtung kann man sagen, dass die Ergebnisse Hypothese (1) im Wesentlichen widerlegen: Die Sprache in der *GEOLino* unterscheidet sich offenbar nicht von der Sprache in der *GEO*. Lediglich die Satzlänge scheint ein distinktives Merkmal zu sein, das sich signifikant in beiden Korpora unterscheidet. In dieser Arbeit habe ich im kleinen Rahmen nachgewiesen, dass meine Fragestellung (1) mit *nein* beantwortet werden muss: Es gibt – bis auf die Satzlänge – keine systematischen sprachlichen Unterschiede zwischen *GEOLino* und *GEO*. Ich habe jedoch eine Reihe von Varianzen beobachten können, die für tendenzielle Unterschiede in den Korpora sprechen und größtenteils in die Richtung der Erwartungen gehen, die von der Literatur vorgegeben werden. Daher könnte

es durchaus sein, dass die von mir betrachteten Merkmale doch relevant sind, die Datenmenge bzw. die Qualität meiner Daten jedoch nicht ausreichend bzw. angemessen ist, um signifikante Unterschiede hervorzubringen. In dieser Form können die Daten die Erwartungen an meine Arbeit jedenfalls nicht erfüllen:

- (1) Die Merkmale scheinen nicht relevant genug zu sein.
- (2) Sie zeigen bis auf eine Ausnahme keine statistisch signifikanten Unterschiede zwischen *GEOlino* und *GEO* an.
- (3) Aufgrund von (2) bieten die Ergebnisse keinen Ausgangspunkt für eine Registerstudie.

Eine Beschreibung der Sprache in der *GEOlino* ist zwar im kleinen Rahmen und nur mit vielen Einschränkungen gelungen, doch ist der Versuch fehlgeschlagen, die *GEOlino* sprachlich nach oben hin zur *GEO* abzugrenzen (mit Ausnahme des Merkmals *Satzlänge*). Aufgrund der Daten kann ich die Fragen (1) und (2) aus Abschnitt 4 nicht mit *ja* beantworten: Die Merkmale, die in der Forschungsliteratur als unangemessen oder untypisch für Sachtexte für junge Leser eingestuft werden, werden in der *GEOlino* im Vergleich zur *GEO* nicht marginal verwendet. Auch werden die Merkmale, die als angemessen oder typisch für Sachtexte für junge Leser eingestuft werden, in der *GEOlino* nicht häufiger verwendet als in der *GEO*.

Die beobachteten Varianzen sprechen jedoch dafür, das Korpus so zu überarbeiten und mit Texten zu erweitern, dass die *SVAutor* und *Geschlecht* ausbalanciert werden können. Eine erneute Analyse wäre dann aufschlussreicher.

7. Ausblick

Nach Lemnitzer/Zinsmeister (2010:52) kann man

„natürlich alle Erkenntnisse, die man durch Beobachtung an einem Korpus gewinnt, allein auf dieses Korpus beziehen. Dies widerspricht aber normalerweise dem Forschungsinteresse der Korpuslinguistik. In der Korpuslinguistik sollen Erkenntnisse gewonnen werden, die über die beobachtete Datenmenge hinaus generalisierbar sind und so unsere Einsicht in die Funktion und Verwendung einer Sprache vertiefen“.

Diesen berechtigten Anspruch kann ich hier noch nicht stellen, da die Datenmenge noch zu klein ist. Diese Arbeit soll ein erstes Herangehen an die Frage sein, ob und wie man die Sprache in der *GEOlino* im Vergleich zur *GEO* beschreiben kann. Dabei bleibt eine ganze Reihe von Fragen bzw. Aspekten offen, die ich hier nicht erschöpfend behandeln kann. Deswegen möchte ich nur kurz die wichtigsten anreißen.

7.1 Optimierung und Erweiterung der Korpora

Wie bereits deutlich wurde, können meine Korpora noch in vielerlei Hinsicht überarbeitet werden. Auf der einen Seite wäre es sinnvoll, sie durch weitere Texte quantitativ zu erweitern, so dass der Einfluss der *SVAutor* und *Geschlecht* kontrolliert werden kann.

Außerdem wäre es sinnvoll, weitere Korpora aus ähnlichen Zeitschriftenreihen wie zum Beispiel *Zeit Wissen/Zeit Leo*, *Spiegel/Dein Spiegel*, *mare/mare aHoi*, *Stern/yuno* oder

P.M./Willi wills wissen zu erstellen. So lässt sich vielleicht für bestimmte sprachliche Merkmale eine ähnliche Frequenz in den Kinderzeitschriften feststellen, die sich von der in den Erwachsenen-Zeitschriften unterscheidet. Damit könnte man die Ergebnisse im Sinne von Lemnitzer/Zinsmeister (ebd.) generalisieren und allgemeine Frequenzmuster sprachlicher Phänomene in Kinderzeitschriften formulieren.

Zudem wäre es angebracht, die bereits annotierten sprachlichen Merkmale noch weiter zu unterteilen; so könnte man zum Beispiel die Wortlänge nicht nur in Buchstaben sondern auch in Silben messen. Damit könnte man Hypothese (2) prüfen und der Frage nachgehen, inwiefern die *GEOLino*-Leser zu einer Zwischengruppe zwischen Leseanfängern und Lesern auf Erwachsenen-Niveau gehören. Die Ergebnisse dieses Vergleichs der Silbenanzahl lassen sich dann in Beziehung zu Deußings (1927) Aufsatzstudie stellen. Im Bereich der Komposition kann man die Annotation der Komplexität der Hauptglieder noch verfeinern, indem man genau bestimmt, welches Glied komplex bzw. simplex ist und auch, welches Wortbildungsverfahren jedem komplexen Glied zugrunde liegt. Im Bereich der adjektivischen Amplifikation ließe sich noch die Unterscheidung treffen zwischen Amplifikationen nach dem Muster des einfachen Vergleichs, wie *stundenlang* und Amplifikationen, die eindeutig eine Steigerung anzeigen, wie *blitzblank*. Im Konjunktionsbereich könnte man Erweiterungen vornehmen, indem man nicht nur adversative und kausale, sondern weitere semantische Klassen annotiert.

Dadurch, dass ich eine Auswahl der sprachlichen Merkmale treffen musste, bleiben wichtige und teilweise auch viel diskutierte Merkmale in meiner Arbeit unberücksichtigt. So wäre es in Anlehnung an Gibson/Levin (1989:261)) interessant, im Bereich der Oberflächenmerkmale nicht nur die absolute Satzlänge zu vergleichen, sondern auch die Länge der Teilsätze. Weiterhin bietet es sich an, die Frequenz folgender Merkmale in *GEOLino* und *GEO* zu vergleichen:

- (1) Erklärung von Fachbegriffen
- (2) Vergleiche, zum Beispiel durch *wie* oder *das heißt*
- (3) Übersetzung von Fremdwörtern bzw. fremdsprachlichen Materials
- (4) Anrede an den Leser durch Imperativ, zum Beispiel *Schaut mal auf die Uhr und zählt mit!*
- (5) umgangssprachliche Reduktionen bzw. Kontraktionen im lexikalischen Bereich, zum Beispiel *Das würd'*; *Los geht's*
- (6) direkte Fragen an den Leser
- (7) Passivkonstruktionen

Dabei lässt sich bezüglich Merkmal (1) und (3) sicher nicht leicht bestimmen, was ein Fach- bzw. Fremdwort ist. Durch die Vielfalt an Erklärungsmöglichkeiten ist auch bei Merkmal (2) die Formulierung einer geeigneten Richtlinie nicht trivial. In diesem Zusammenhang wäre auch ein Vergleich der Frequenz von Metaphern in *GEOLino* und *GEO* interessant. Merkmale (4) bis (7) ließen sich hingegen einfacher annotieren. So sind Imperativ-Konstruktionen auf der POS-Ebene mit *VVIMP* und *VAIMP* vorannotiert; Reduktionen bzw. Kontraktionen ließen sich vielleicht am Vorkommen des Apostrophs festmachen; direkte Fragen an den Leser könnte man anhand von Fragezeichen manuell annotieren und Passivkonstruktionen lassen sich vielleicht durch Vorkommen von *werden*

aufspüren. Für die Relevanz dieser zusätzlichen Merkmale gibt es zahlreiche Hinweise in der Literatur.

7.2 Registerfrage

Neben der Erweiterung meiner Korpora gibt es viele offene Aspekte, die ich nicht oder nicht ausreichend behandeln konnte. So bleibt zum Beispiel Fragestellung (2) offen: Handelt es sich bei *GEOlino* und *GEO* um verschiedene Register? Die Ergebnisse bieten keinen Anhaltspunkt dafür, insbesondere, da ich bezüglich der Verwendung von attributiven Adjektiven und der Verwendung von direkter Anrede an den Leser keine signifikanten Unterschiede feststellen konnte. Lediglich für letzteres Merkmal konnte ich einen tendenziellen Unterschied erkennen. Jedoch muss man sich hier die Frage stellen, ob die von mir betrachteten Merkmale für eine Registeranalyse überhaupt sinnvolle Parameter sind? Wenn ich untersuchen möchte, ob ein bestimmtes Register vorliegt, würde ich immer die Frage stellen, ob man bezüglich des betrachteten Parameters Ausdrucksalternativen hat. Erst wenn ich diese Frage mit *ja* beantworten kann, ist der Parameter auch sinnvoll angesetzt. Dies ist jedoch bei meinen Merkmalen nicht der Fall: Welche Ausdrucksalternativen gibt es zu nominalen Abstrakta oder adversativen bzw. kausalen Konjunktionen? Daher nehme ich an, dass sich die untersuchten Merkmale des univariaten Vergleichs nicht in gleichem Maße für eine Registeranalyse eignen.

Bezogen auf eine Registeranalyse in *GEOlino* und *GEO* bedeutet Hypothese (1), dass die Variablen im Register *GEOlino* andere Varianten aufweisen als im Register *GEO*. Hier müsste man zunächst einmal die Variablen ermitteln, bevor man die Unterschiede in deren Ausprägungen herausarbeiten kann.

Es ist auch fraglich, ob man die Frage nach einem *GEOlino*-Register so pauschal stellen kann. Wahrscheinlich gibt es innerhalb der *GEOlino* schon verschiedene Register, daher müsste man zunächst prüfen, ob und welche Register *GEOlino*-intern vertreten sind. Sind verschiedene Register vertreten, muss die Frage nach dem *GEOlino*-Register verworfen werden, da man nicht mehrere Einzelregister zusammenfassen kann. Man müsste prüfen, ob es zu einem *GEOlino*-Einzelregister ein entsprechendes Register in der *GEO* gibt und kann dann gegebenenfalls diese Einzelregister miteinander vergleichen, um *GEOlino*-typische sprachliche Charakteristiken herauszuarbeiten.

Was in meiner Arbeit auch offen bleiben muss, ist die sprachliche Abgrenzung der *GEOlino* nach unten hin zur *GEOMini*. Erst mit dieser zweiten Abgrenzung könnte man feststellen, ob es eine *GEOlino*-typische sprachliche Charakteristik gibt, die weder in der *GEO* noch in der *GEOMini* zu finden ist.

7.3 Doppeltexte

Einen weiteren interessanten Aspekt habe ich bereits in Abschnitt 4.1.3 angesprochen. Es geht hier um Autoren, die sowohl für die *GEOlino* als auch für die *GEO* schreiben und zwar über das gleiche Thema. Teilweise sind hier auch Titel und Bilder in den Artikeln identisch. Diese Texte unterscheiden sich jeweils tatsächlich nur darin, für welches Publikum sie verfasst wurden und diese Unterschiede werden an der Wahl der sprachlichen Mittel deutlich. Unterschiede zwischen diesen Texten können weder auf den Einfluss der SV *Autor* noch der SV *Geschlecht* zurückgeführt werden, sondern sind der UV *Korpus* zuzuschreiben. Hier könnte man also in einem sehr kleinen Rahmen (nur ein einziger

Autor) die Analyse durchführen und feststellen, inwiefern die Daten die Erwartungen bezüglich Hypothese (1) widerspiegeln. Dass es Unterschiede gibt, kann man bereits erkennen, wenn man sich die Verteilung der POS-tags in *GEOlino5* und *GEO5* ansieht (beide sind vom selben Autor verfasst): Die Frequenzen unterscheiden sich bezüglich jedes einzelnen POS-tags und der Chi²-Test zeigt, dass der Unterschied signifikant ist (Chi²= 25,2, df= 14, p= 0,03). Entsprechendes gilt auch für *GEOlino2* und *GEO46* (Chi²= 24,2, df= 14, p= 0,04). Ich betrachte hier POS-tags, die mindestens 50 Mal über die beiden Texte des jeweiligen Autors summiert auftreten.

7.4 Aspekt *Verständlichkeit*

Die große Frage nach der Verständlichkeit der Texte in der *GEOlino* konnte ich hier noch nicht behandeln. Verständlichkeit ist jedoch ein wichtiger Faktor bei der Betrachtung von Texten für junge Leser. Man muss sich fragen, inwiefern die *GEOlino* ihr Publikum sprachlich erreicht: Sind die Artikel angemessen geschrieben, so dass die jungen Leser den Inhalt aufnehmen und verstehen können? Würden sie vielleicht Texte, die für Erwachsene geschrieben sind, gleich gut verstehen und wenn nicht, bei welchen Konstruktionen treten Verständnisprobleme auf? Die überdachende Frage, was angemessene Sprache im Alter von 10 bis 12 Jahren eigentlich ausmacht, welche sprachlichen Kriterien also Verständlichkeit bestimmen, soll und kann hier noch nicht zuverlässig beantwortet werden und ist das langfristige Ziel weiterer Arbeit auf diesem Gebiet.

8. Literaturverzeichnis

- Adamzik, Kirsten (2008): Textsorten und ihre Beschreibung. In: Janich, Nina (Hrsg.): Textlinguistik. 15 Einführungen. Narr Studienbücher, Tübingen. S. 145-176.
- Antos, Gerd (2008): Schriftliche Textproduktion. Formulieren als Problemlösung. In: Janich, Nina (Hrsg.): Textlinguistik. 15 Einführungen. Narr Studienbücher, Tübingen. S. 237-254.
- Baacke, Dieter; Lauffer, Jürgen (1994): Nicht nur schöner Schein. Kinder- und Jugendzeitschriften in Deutschland. Übersicht und Empfehlungen. GMK (Medienpädagogische Handreichungen 4) (Schriften zur Medienpädagogik 16). GMK, Bielefeld.
- Berthold, Sabine (2010): Ökonomie der Wissensgesellschaft. Wirtschaft als Sujet in der Sachliteratur für Kinder und Jugendliche. In: Meissner, Almuth; Oels, David (Hrsg.): Sachtexte für Kinder und Jugendliche. Reihe *Non Fiktion* (5, 1/2). Wehrhahn Verlag, Hannover. S. 45-56.
- Baurmann, Jürgen (2009): Sachtexte lesen und verstehen: Grundlagen, Ergebnisse, Vorschläge für einen kompetenzfördernden Unterricht. Klett/Kallmeyer, Stuttgart.
- Biber, Douglas (2006): University Language. A corpus-based study of spoken and written registers. Reihe *Studies in Corpus Linguistics* (23). John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Biber, Douglas (2009): Multi-dimensional Approaches. In: Lüdeling, Anke; Kytö, Merja (Hrsg.): Corpus Linguistics. An International Handbook. Reihe *Handbücher zur Sprach- und Kommunikationswissenschaft* (29). De Gruyter, Berlin/New York. S. 822-855.
- Biber, Douglas (1995): Dimensions of register variation: a cross-linguistic comparison. Cambridge University Press, Cambridge.
- Biber, Douglas; Conrad, Susan (2009): Register, genre, and style. Cambridge University Press, Cambridge.
- Biber, Douglas; Conrad, Susan; Reppen, Randi (2002): Corpus linguistics. Investigating Language Structure and Use. Cambridge University Press, Cambridge.
- Bos, Wilfried; Lankes, Eva-Maria; Prenzel, Manfred; Schwippert, Knut; Walther, Gerd; Valtin, Renate (Hrsg.) (2003): Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich. Waxmann, Münster/New York/München/Berlin.
- Bredel, Ursula; Fuhrhop, Nanna; Noack, Christina (2011): Wie Kinder lesen und schreiben lernen. Francke, Tübingen.
- Buscha, Joachim (1989): Lexikon deutscher Konjunktionen. VEB Verlag Enzyklopädie, Leipzig.
- Bußmann, Hadumod (Hrsg.) (2002): Lexikon der Sprachwissenschaft. Kröner, Stuttgart.
- Christmann, Ursula; Groeben, Norbert (2006): Anforderungen und Einflussfaktoren bei Sach- und Informationstexten. In: Groeben, Norbert; Hurrelmann, Bettina (Hrsg.): Lesekompetenz. Bedingungen, Dimensionen, Funktionen. Juventa, Weinheim/München. S. 150-173.
- Clark, Eve V. (2009): Lexical meaning. In: Bavin, Edith L. (Hrsg.): The Cambridge Handbook of child Language. Cambridge University Press, Cambridge. S. 283-299.
- Deußing, Hans (1927): Der sprachliche Ausdruck des Schulkindes. Statistische und experimentelle Untersuchungen zur Entwicklungspsychologie der Sprache. In: Helmers, Hermann (Hrsg.) (1969): Zur Sprache des Kindes. Wiss. Buchges., Darmstadt. S. 60-131.
- Diedrichs, Ulf (2010): Annäherungen an das Sachbuch. Zur Geschichte und Definition eines umstrittenen Begriffs. Reihe *Arbeitsblätter für die Sachbuchforschung* (18). *Historische Reihe* (6). Humboldt-Universität zu Berlin/Universität Hildesheim, Berlin/Hildesheim.
- Dietrich, Rainer (2007): Psycholinguistik. Metzler, Stuttgart/Weimar.
- Dijkstra, Ton; Kempen, Gerard (1993): Einführung in die Psycholinguistik. Verlag Hans Huber, Bern/Göttingen/Toronto/Seattle.
- Doderer, Klaus (1960): Zur sprachlichen Situation der Schulanfänger und der Grundschulkinder. In: Helmers, Hermann (Hrsg.) (1969): Zur Sprache des Kindes. Wiss. Buchges., Darmstadt. S. 308-325.

- DuBay, William (2004): The Principles of Readability. [<http://www.impact-information.com/impactinfo/readability02.pdf>] vom 12.7.2011.
- Ehri, Linnea C. (2005): Learning to Read Words: Theory, Findings, and Issues. In: Cain, Kate (Hrsg.): *Scientific Studies of Reading* (9/2). Erlbaum, Mahwah/New Jersey. S. 167-188.
- Eisenberg, Peter (2004): *Grundriß der deutschen Grammatik*. Bd. 1: Das Wort. Metzler, Stuttgart/Weimar.
- Feilke, Helmuth (1996): From Syntactical to Textual Strategies of Argumentation. Syntactical Development in Written Argumentative Texts by Students Aged 10 to 22. In: Golder, Caroline (Hrsg.): *Argumentation in Texts: Developmental Studies in Production and Comprehension* (10/2). Kluwer, Dordrecht. S. 197-212.
- Ferguson, Charles A. (1994): Dialect, Register, and Genre: Working Assumptions About Conventionalization. In: Biber, Douglas; Finegan, Edward (Hrsg.): *Sociolinguistic perspectives on register*. Oxford University Press, New York/Oxford. S.15-30.
- GEOLino-/GEO-Korpora*: <http://www.sfb632.uni-potsdam.de/d1/annis/download.html> (in Kürze einsehbar).
- Gibson, Eleanor J.; Levin, Harry (1989): *Die Psychologie des Lesens*. Fischer Taschenbuch Verlag, Frankfurt a.M.
- Gierlich, Heinz (2005): Sachtexte als Gegenstand des Deutschunterrichts – einige grundsätzliche Überlegungen. In: Fix, Martin; Jost, Roland (Hrsg.): *Sachtexte im Deutschunterricht: für Karlheinz Fingerhut zum 65. Geburtstag*. Schneider-Verlag Hohengehren, Baltmannsweiler. S. 25-46.
- Göpferich, Susanne (2008): Textverstehen und Textverständlichkeit. In: Janich, Nina (Hrsg.): *Textlinguistik. 15 Einführungen*. Narr Studienbücher, Tübingen. S. 291-312.
- Groeben, Norbert (1982): *Leserpsychologie: Textverständnis – Textverständlichkeit*. Aschendorff, Münster.
- Hiller, Florian (2010): *Sachtexte erschließen. Eine empirische Studie zur Förderung der Lesekompetenz*. Fillibach, Freiburg im Breisgau.
- Höhle, Barbara (2010): Erstspracherwerb: Wie kommt das Kind zur Sprache? In: Höhle, Barbara (2010) (Hrsg.): *Psycholinguistik*. Akademie Verlag, Berlin.
- Hummelsberger, Siegfried (2003): ‚Sachtext-Leser‘ oder ‚Sach-Bearbeiter‘? Wie sachdienlich ist die Rede von Sachtexten in eigener Sache? In: Abraham, Ulf; Bremerich-Vos, Albert; Frederking, Volker; Wieler, Petra (Hrsg.): *Deutschdidaktik und Deutschunterricht nach PISA*. Fillibach, Freiburg. S. 330-346.
- Hunston, Susan (2008): Collecting strategies and design decisions. In: Lüdeling, Anke; Kytö, Merja (Hrsg.): *Corpus Linguistics. An International Handbook. Reihe Handbücher zur Sprach- und Kommunikationswissenschaft* (29). De Gruyter, Berlin/New York, S. 154-168.
- Johnson, Janice; Pascual-Leone, Juan (1989): Developmental Levels of Processing in Metaphor Interpretation. In: Bjorklund, David F. (Hrsg.): *Journal of Experimental Child Psychology* (48/1). S.1-31.
- Klosa, Annette; Scholze-Stubenrecht, Werner; Wermke, Matthias (Hrsg.): *Duden*. Bd. 4: *Grammatik der deutschen Gegenwartssprache*. Dudenverlag, Mannheim/Leipzig/Wien/Zürich.
- Kommerell, Kathrin (2008): *Journalismus für junge Leser*. UVK Medien, Konstanz.
- Kramer, Undine (1990): *Sachtexte und Rezeption. Untersuchungen zur prototypischen Beschreibung ausgewählter Sachtexte und ihrer Rezeption im Muttersprachunterricht*. Dissertation. Humboldt-Universität zu Berlin, Berlin.
- Kühn, Julia (2010): Nur kein "Babykram". In: Meissner, Almuth; Oels, David (Hrsg.): *Sachtexte für Kinder und Jugendliche. Reihe Non Fiktion* (5,1/2). Wehrhahn Verlag, Hannover. S. 129-133.
- Kurz, Ellen (2000): *Kinderzeitschriften. Erstellung einer Marktübersicht sowie einer annotierten Empfehlungsliste im Internet*. Diplomarbeit im Fach Kinder- und Jugendmedien Studiengang Öffentliches Bibliothekswesen an der Fachhochschule Stuttgart – Hochschule für Bibliotheks- und Informationswesen.

- Langer, Inghard; Schulz v. Thun, Friedemann; Tausch, Reinhard (2011): *Sich verständlich ausdrücken*. Ernst Reinhardt Verlag, München/Basel.
- Lemnitzer, Lothar; Zinsmeister, Heike (2010): *Korpuslinguistik. Eine Einführung*. Narr Studienbücher, Tübingen.
- Lust, Barbara (2009): *Child Language. Acquisition and Growth*. Cambridge University Press, Cambridge.
- May, Peter (1996): *Erfassung bildnerischer und textueller Gestaltungsfähigkeiten in der Schule. Hamburger Leitfaden für die Bewertung von Bild- und Textprodukten (HLBT)*. Mitarbeit v. Annette Juchems. [http://www1.uni-hamburg.de/psycholo/frames/projekte/PLUS/PLUS_doc/HLBT_Manual.pdf] vom 12.7.2011.
- Meibauer, Jörg (2008): *Pragmatik. Eine Einführung*. Reihe *Stauffenburg-Einführungen (12)*. Stauffenburg-Verlag, Tübingen.
- Meindl, Claudia (2011): *Methodik für Linguisten. Eine Einführung in Statistik und Versuchsplanung*. Narr Studienbücher, Tübingen.
- Meissner, Almuth (2010): *Nach PISA ist vor PISA. Sachtexte lesen im Deutschunterricht*. In: Meissner, Almuth; Oels, David (Hrsg.): *Sachtexte für Kinder und Jugendliche*. Reihe *Non Fiktion (5,1/2)*. Wehrhahn Verlag, Hannover. S. 73-87.
- Milone, Michael (2008): *The Development of ATOS: The Renaissance Readability Formula*. [<http://doc.renlearn.com/KMNet/R004250827GJ11C4.pdf>] vom 12.7.2011.
- Moers, Edelgard (2004): *Informierendes Lesen. Mit Informationen aktiv umgehen; Lesekompetenz stärken; Beispiele aus der Praxis*. Cornelsen Scriptor, Berlin.
- Niederhauser, Jürg (1999): *Wissenschaftssprache und populärwissenschaftliche Vermittlung*. Reihe *Forum für Fachsprachen-Forschung (53)*. Narr Studienbücher, Tübingen.
- Nussbaumer, Markus; Sieber, Peter (1994): *Texte analysieren mit dem Zürcher Textanalyseraster*. In: Sieber, Peter (Hrsg.): *Sprachfähigkeiten - besser als ihr Ruf und nötiger denn je! Ergebnisse und Folgerungen aus einem Forschungsprojekt*. Reihe *Sprachlandschaft (12)*. Sauerländer, Aarau/Frankfurt a.M./Salzburg. S.142-185.
- Pech, Klaus-Ulrich (2010): *Grenzenloser Belehrungsoptimismus. Sachbücher für junge Leser im 19. Jahrhundert*. In: Meissner, Almuth; Oels, David (Hrsg.): *Sachtexte für Kinder und Jugendliche*. Reihe *Non Fiktion (5,1/2)*. Wehrhahn Verlag, Hannover. S. 11-27.
- Pelz, Heidrun (2004): *Linguistik. Eine Einführung*. Hoffmann und Campe, Hamburg.
- Rayner, Keith; Pollatsek, Alexander (1989): *The psychology of reading*. Erlbaum, Hillsdale/New Jersey.
- Reitsma, Pieter; Siegel, Linda S.; Verhoeven, Ludo (2011): *Cognitive and linguistic factors in reading acquisition*. In: Joshi, R. Malatesha (Hrsg.): *Reading and Writing: An interdisciplinary journal (24/4)*. Springer, Dordrecht. S. 387-394.
- Rosebrock, Cornelia; Nix, Daniel (2008): *Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung*. Schneider-Verlag Hohengehren, Baltmannsweiler.
- Scherer, Carmen (2006): *Korpuslinguistik*. Reihe *Kurze Einführungen in die germanistische Linguistik*, Bd. 2. Winter, Heidelberg.
- Sommer, Michael (1994): *Die Kinderpresse in der Bundesrepublik Deutschland. Angebot, Konzepte, Formen, Inhalte*. Kovač, Hamburg.
- Steinlein, Rüdiger (2010): *Fiktionalität und Nicht-Fiktionalität im Kinder- und Jugendsachbuch. Überlegungen zur Bestimmung eines vielgestaltigen Genres*. In: Meissner, Almuth; Oels, David (Hrsg.): *Sachtexte für Kinder und Jugendliche*. Reihe *Non Fiktion (5,1/2)*. Wehrhahn Verlag, Hannover. S. 29-44.
- Tracy, Rosemarie (2008): *Wie Kinder Sprachen lernen. Und wie wir sie dabei unterstützen können*. Francke, Tübingen.

Willenberg, Heiner (2005): Ein handhabbares System, um Textschwierigkeiten einzuschätzen. Vorschläge für eine Textdatenbank von Sachtexten. In: Fix, Martin; Jost, Roland (Hrsg.): Sachtexte im Deutschunterricht: für Karlheinz Fingerhut zum 65. Geburtstag. Schneider-Verlag Hohengehren, Baltmannsweiler. S. 94-106.

Wittwen, Andreas (1995): Infotainment. Fernsehnachrichten zwischen Information und Unterhaltung. Reihe *Zürcher germanistische Studien (43)*. Lang, Bern/Berlin/Frankfurt a.M./New York/Paris/Wien.

A. Richtlinien

Name	Ebene	Motivation aus Literatur	Annotation	Tags	Beispiel
Wortlänge	WL	<ul style="list-style-type: none"> - Wort- und Satzlänge beeinflussen die Lesbarkeit und das Verstehen eines Textes - Kurze Wörter und Sätze für junge Leser bevorzugen 	Wortlänge in Buchstaben		
Satzlänge	[Berechnung direkt in <i>R</i>]		Satzlänge in Wörtern zwischen Satzendzeichen		
nominale Komposita	nominales.kompositum	<ul style="list-style-type: none"> - Hochfrequente Verwendung von Komposita wirkt sich negativ auf das Verstehen aus - Komposition verlängert die Wörter (siehe Merkmal <i>Wortlänge</i>) - Die meisten Komposita im Deutschen sind nominal 	nicht-lexikalisierte Substantive (POS- <i>tagNN</i> + <i>TRUNC</i>)	[nk]	Raumanzug Vier-Tage-Bart
			Zusammengezogene Komposita werden mehrfach annotiert	2x [nk]	Brems- und Steuerrakete
			lexikalisierte Substantive (POS- <i>tagNN</i> + <i>TRUNC</i>)	[nonk]	Bahnhof Wasserstoff
			Nicht: Reine Eigennamen (POS- <i>tagNE</i>); fremdsprachliches Material		
Anzahl Hauptglieder	anzahl.glieder	<ul style="list-style-type: none"> - Nominale Komposita sind theoretisch unendlich erweiterbar (siehe Merkmal <i>Wortlänge</i>) 	Hauptglieder von [nk] numerisch	[2] [3] ...	Raumanzug [2] Vier-Tage-Bart [2]
			Zusammengezogene Komposita werden mehrfach annotiert	2x [2]	Brems- und Steuerrakete
Komplexität Hauptglieder	komplexitaet.glieder	<ul style="list-style-type: none"> - Hauptglieder von Komposita können in sich komplex aufgebaut sein - kognitiven Arbeitsaufwand für junge Leser eher gering halten 	[nk] mit mindestens einem komplexen Hauptglied	[ko]	Uhrenrädchen
			[nk] mit keinem komplexen Hauptglied	[si]	Wollfänger Wartehalle

Name	Ebene	Motivation aus Literatur	Annotation	Tags	Beispiel
Anzahl Adjektive	[automatisch auf der POS-Ebene vorannotiert]	<ul style="list-style-type: none"> - Lesbarkeitsforschung: für junge Leser viele Adjektive verwenden - Journalismus: für junge Leser wenig Adjektive verwenden 	attributive Adjektive	[ADJA]	in <i>naher</i> Zukunft
			adverbiale oder prädikative Adjektive	[ADJD]	wie <i>schwierig</i> es ist, im Weltraum zu arbeiten
Verhältnis Adj _{attr} zu Adj _{präd}	[automatisch auf der POS-Ebene vorannotiert]	<ul style="list-style-type: none"> - Adj_{attr} werden im Laufe der Sprachentwicklung immer wichtiger - Verwendungshäufigkeit von Adj_{attr} und Adj_{präd} ist Register unterscheidend 	attributive Adjektive	[ADJA]	<i>extreme</i> Bedingungen
			adverbiale oder prädikative Adjektive	[ADJD]	<i>Schwerelos</i> sind auch Dinge, die man in die Luft wirft
Komparation	komparation	<ul style="list-style-type: none"> - Journalismus: keine Superlative für junge Leser verwenden - Textlinguistik/Sachtextforschung: Komparation spielt wesentliche Rolle in der Wissensvermittlung 	Positiv (POS-tagADJA, ADJD, TRUNC)	[pos]	Bist du <i>irre</i> ?
			Komparativ (POS-tagADJA, ADJD, TRUNC)	[komp]	<i>leichter</i> als federleicht
			Superlativ (POS-tagADJA, ADJD, TRUNC)	[sup]	der <i>kleinste</i> Fehler
			Nicht: Partizipien, die nicht eindeutig als Adjektiv lexikalisiert sind; Absolutadjektive; Homographen; fremdsprachliches Material		

Name	Ebene	Motivation aus Literatur	Annotation	Tags	Beispiel
adjektivische Amplifikationen	amplifikation	<ul style="list-style-type: none"> - für junge Leser anschaulich schreiben - eigene Herleitung: Amplifikationen sind besonders anschaulich 	Adjektive (POS-tagADJA, ADJD, TRUNC), die einen wortinhärenten Vergleich aufweisen, der als Komparation verstanden werden kann	[ampl]	blitzblank federleicht
			Adjektive, die einen wortinhärenten Vergleich aufweisen, der nicht als Komparation verstanden werden kann	[noampl]	schubladengroß
			Nicht: fremdsprachliches Material		
nominale Abstrakta	nominales.abstraktum	<ul style="list-style-type: none"> - Abstrakta beeinflussen das Textverständnis - wirken sich Verständnishemmend auf junge Leser aus - möglichst keine Abstrakta für junge Leser verwenden 	Substantive (POS-tagNN + TRUNC) mit der Endung <i>heit, ie [i:], ik, ion, ismus, ität, keit, nz, tur</i> oder <i>ung</i>	[abst]	Weltraumstation Montur Hochspannung
			alle restlichen Substantive (POS-tagNN + TRUNC), dabei: relevante Endung am Kompositions-Erstglied; Eigennamen mit hohem Lexikalisierungsstatus	[noabst]	Raumanzug Galaxieströme Internationale Raumstation
			Nicht: reine Eigennamen (POS-tagNE); fremdsprachliches Material		

Name	Ebene	Motivation aus Literatur	Annotation	Tags	Beispiel
Anzahl Konjunktionen	[automatisch auf der POS-Ebene vorannotiert]	<ul style="list-style-type: none"> - Lesbarkeitsforschung: Konjunktionen wirken Verständnis fördernd 	subordinierende Konjunktionen	[KOU1] [KOUS]	<i>um</i> ein Gefühl für sie zu bekommen <i>Wenn</i> es nach ihm ginge, ...

Name	Ebene	Motivation aus Literatur	Annotation	Tags	Beispiel
		- Lesbarkeitsforschung: Konjunktionen sind potenziell Verständnis hemmend	koordinierende Konjunktionen	[KON]	<i>Doch</i> da gibt es ein Problem
			Vergleichskonjunktionen	[KOKOM]	Es sieht aus <i>wie</i> eine gigantische Getränkedose
adversative Konjunktionen	konjunktion.semantisch	<ul style="list-style-type: none"> - verleihen einem Text Explizitt und Kohrenz - Sachtextforschung: wirken Verstndnis frdernd - Psycholinguistik: Explizitt fr junge Leser nicht wichtiger als fr Erwachsene 	Konjunktionen <i>aber, allein, (je)doch, sondern, whrend, wohingegen</i> jeweils in adversativer Bedeutung	[advers]	<i>Doch</i> dann wurde die Reise abgesagt
kausale Konjunktionen			Konjunktionen <i>um so/desto ..., als, da, denn, um so mehr/weniger als, weil, zumal (da)</i> jeweils in kausaler Bedeutung	[kaus]	<i>Denn</i> wer im Columbus-Modul forschen will
Leserkontakt	anrede	<ul style="list-style-type: none"> - direkte Anrede steigert Lesemotivation und Verstndnis bei jungen Lesern - Verwendung von Personalpronomen der 2. Person in geschriebenen Informationstexten ist Register unterscheidend 	Personalpronomen 2. Pers. (<i>du/ihr, Sie</i>) (POS-tagPPER)	[anr]	Schafft <i>ihr</i> es, an den winzigen Uhrenrdchen zu drehen?
			restliche Personalpronomen (POS-tagPPER)	[noanr]	<i>Sie</i> steht blitzblank geputzt in der berhmtesten Wartehalle der Welt

Name	Ebene	Motivation aus Literatur	Annotation	Tags	Beispiel
Metasprachliches	titel	- eigene berlegung: metasprachliche Elemente haben eine andere sprachliche Charakteristik als Flietext	berschriften	[t]	
	untertitel		Unterberschriften	[ut]	
	teaser		Teaser	[tea]	

B. Lexikalisierte Komposita

Achterbahn	Gegenteil	Paraterraforming	Vollmond
Astrophysik(er)	Gleichgewicht	Quadratkilometer	Volltreffer
Atmosphäre	Großteil	Regenbogen	Wagemut
Ausmaß	Gutachter	Rucksack	Wasserstoff
Bahnhof	Halbinsel	Sauerstoff	Weihnacht(en)
Beispiel	Handbuch	Scheiterhaufen	Weltall
Bildschirm	Handschuhe	Schublade	Weltraum
Bösewicht	Hauptsache	Schwerkraft	Werkstoff
Breitengrad	Hemisphäre	Science-Fiction	Werkzeug
Bullauge	Hinsicht	Sehnsucht	Wettrennen
Check-Out	Höhepunkt	Sonnenfinsternis	Wisconsin-Madison
Cockpit	Jahrhundert	Spaziergang	Wochenende
Countdown	Kahlschlag	Spitzname	Zeitfenster
Durchmesser	Kohlendioxid	Sternschnuppe	Zeitlupe
Echolot	Kohlenstoff	Stickstoff	Zeitpunkt
Edelgas	Kohlenwasserstoff	Stoffwechsel	Zeitraffer
England	Kreislauf	Stonehenge	Zeitraum
Ernstfall	Kunststoff	Störenfried	Zwischenfall
Fahrstuhl	Lebewesen	Streichholz	
Fahrzeug	Lichtjahr	Supernova	Hardware (engl.)
Fallschirm	Luftdruck	Tankstelle	Software (engl.)
Feiertag	Mikrowelle(n)	Telefonzelle	Workshop (engl.)
Fernbedienung	Milchstraße	Treibhaus	
Fernglas	Mittelmeer	Treibstoff	
Fernrohr	Mittelpunkt	Triebwerk	
Feuerwerk	Mondfinsternis	Umwelt	
Fliehkraft	Nachsicht	Unruhe	
Flughafen	Neumond	Unterschrift	
Flugzeug	Notfall	Ursprung	
Fußboden	Oberfläche	Vollgas	

C. Grafiken zum univariaten Vergleich

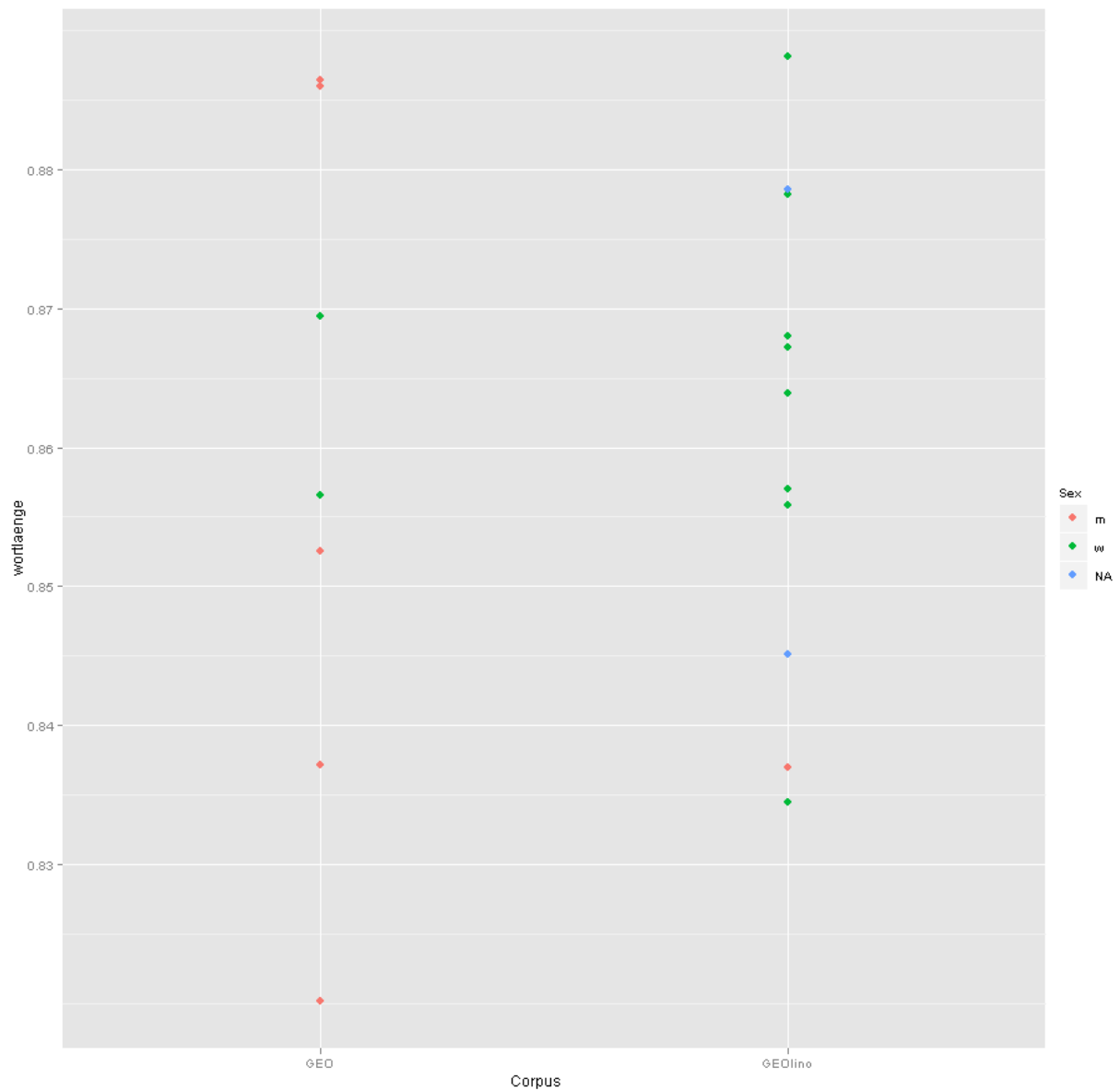


Abbildung 6: Durchschnittliche Wortlänge

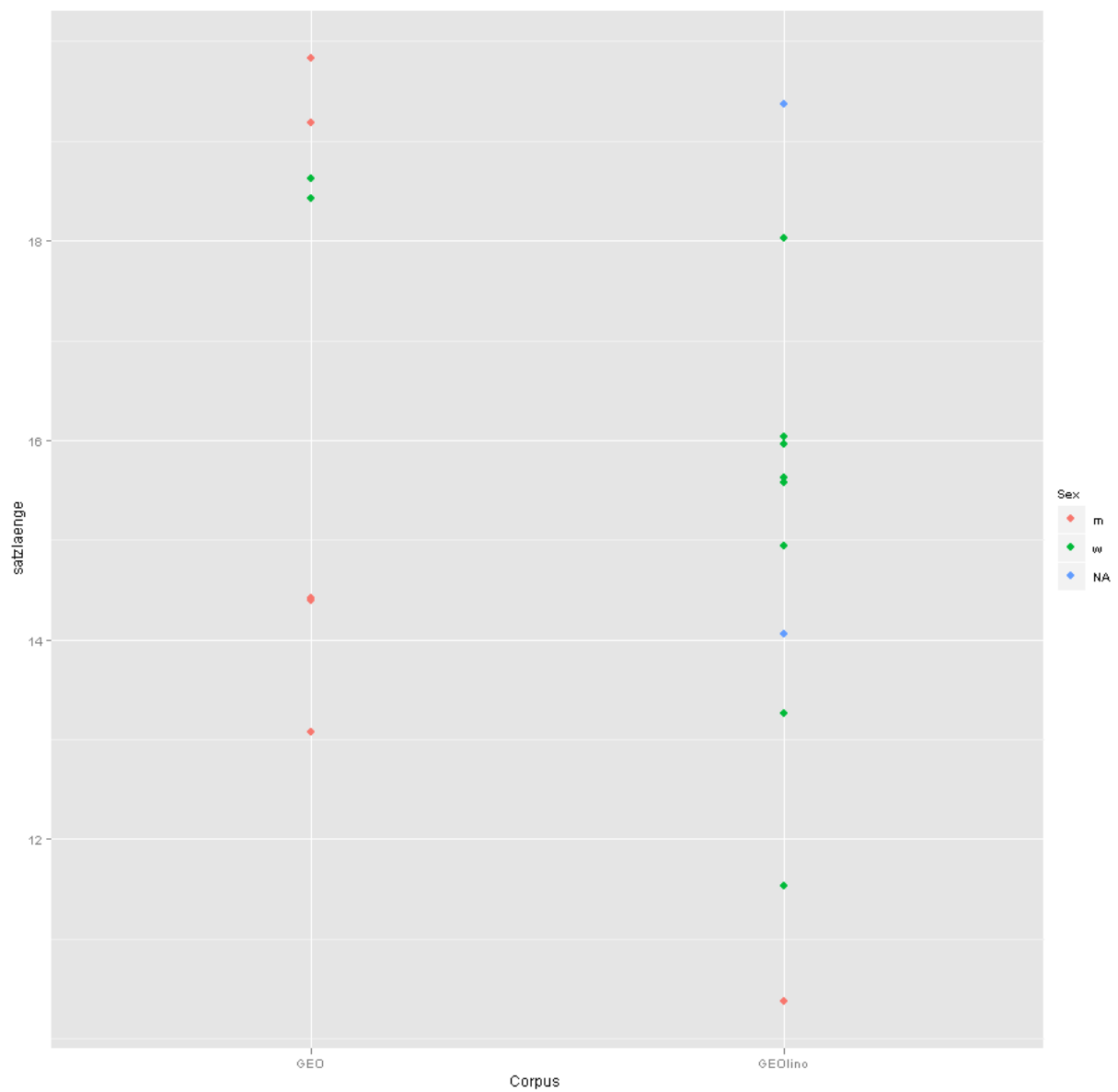


Abbildung 7: Durchschnittliche Satzlänge

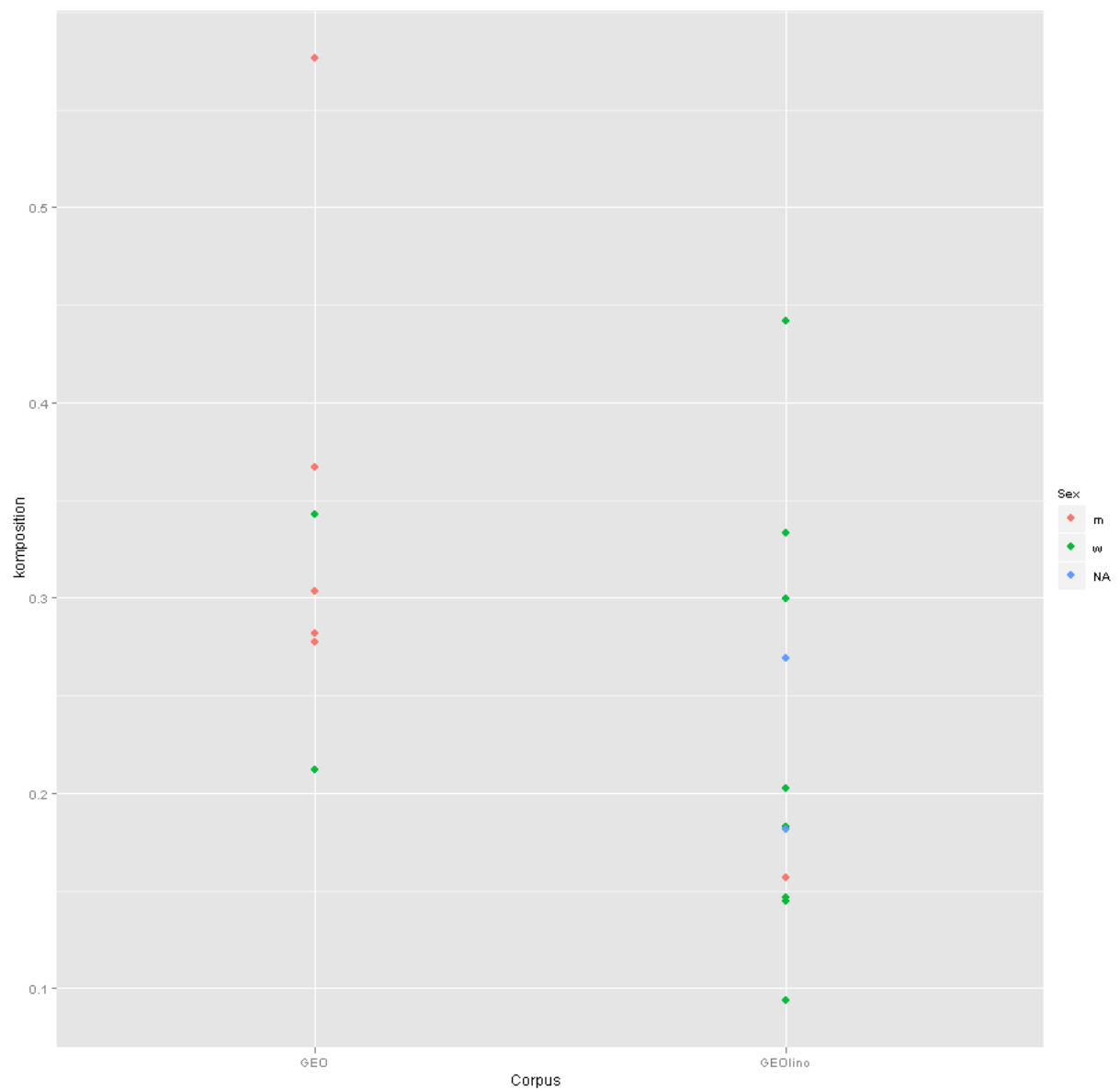


Abbildung 8: Verhältnis von Komposita zu Nicht-Komposita

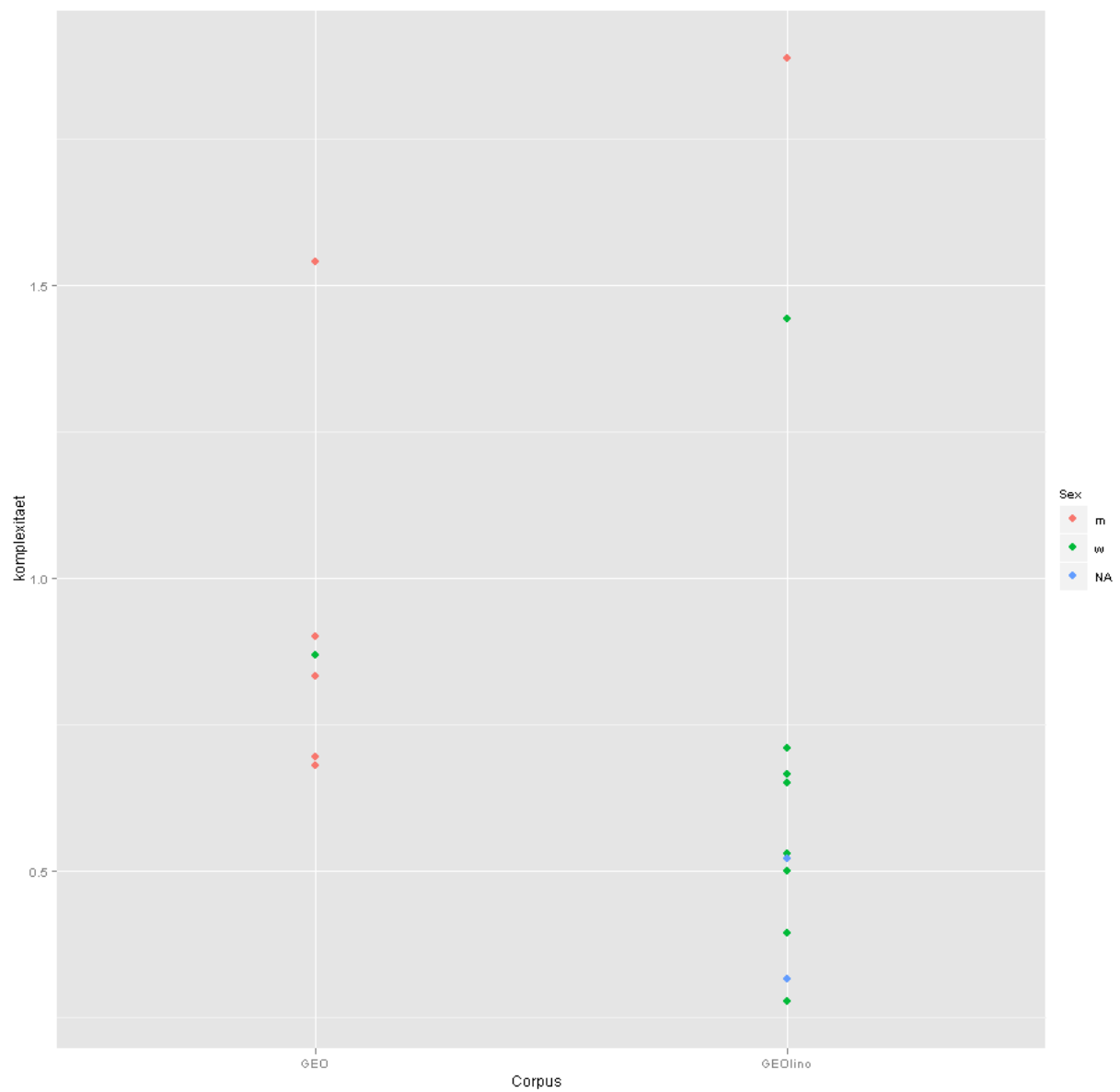


Abbildung 9: Verhältnis von komplexen zu simplex Kompositions-Hauptgliedern

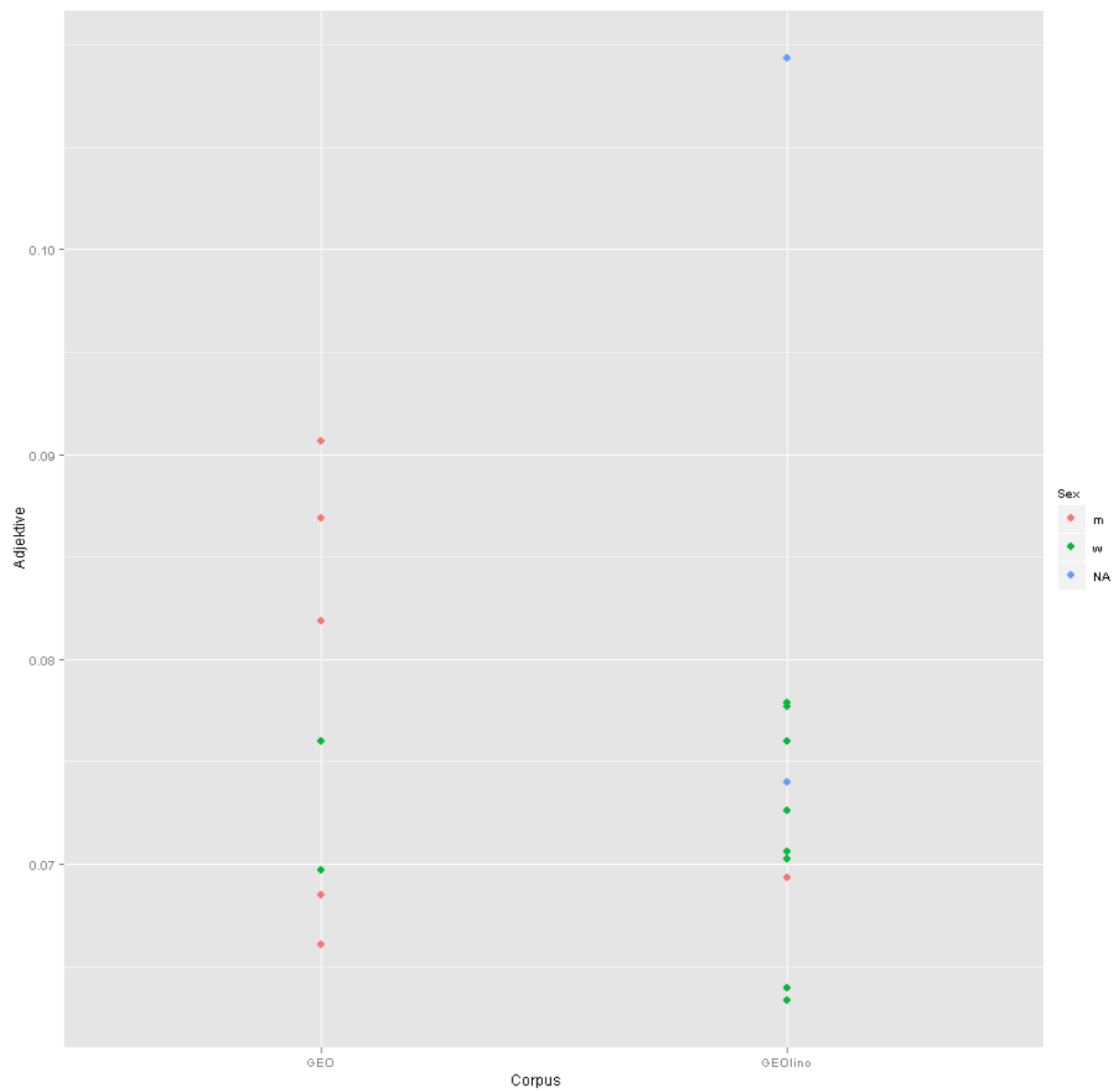


Abbildung 10: Anteil der Adjektive an der Gesamt-Tokenzahl

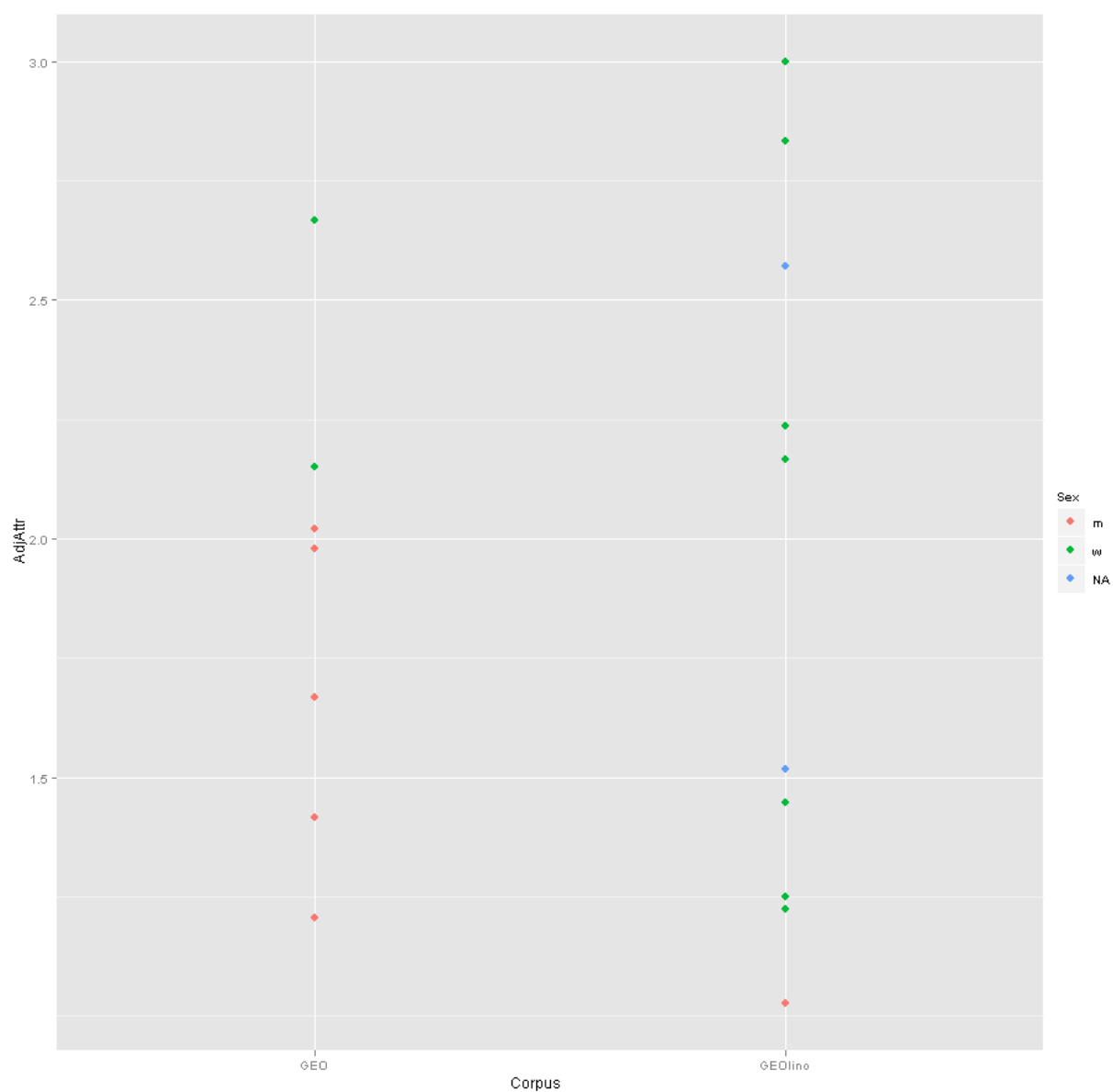


Abbildung 11: Verhältnis von attributiven zu prädikativen Adjektiven

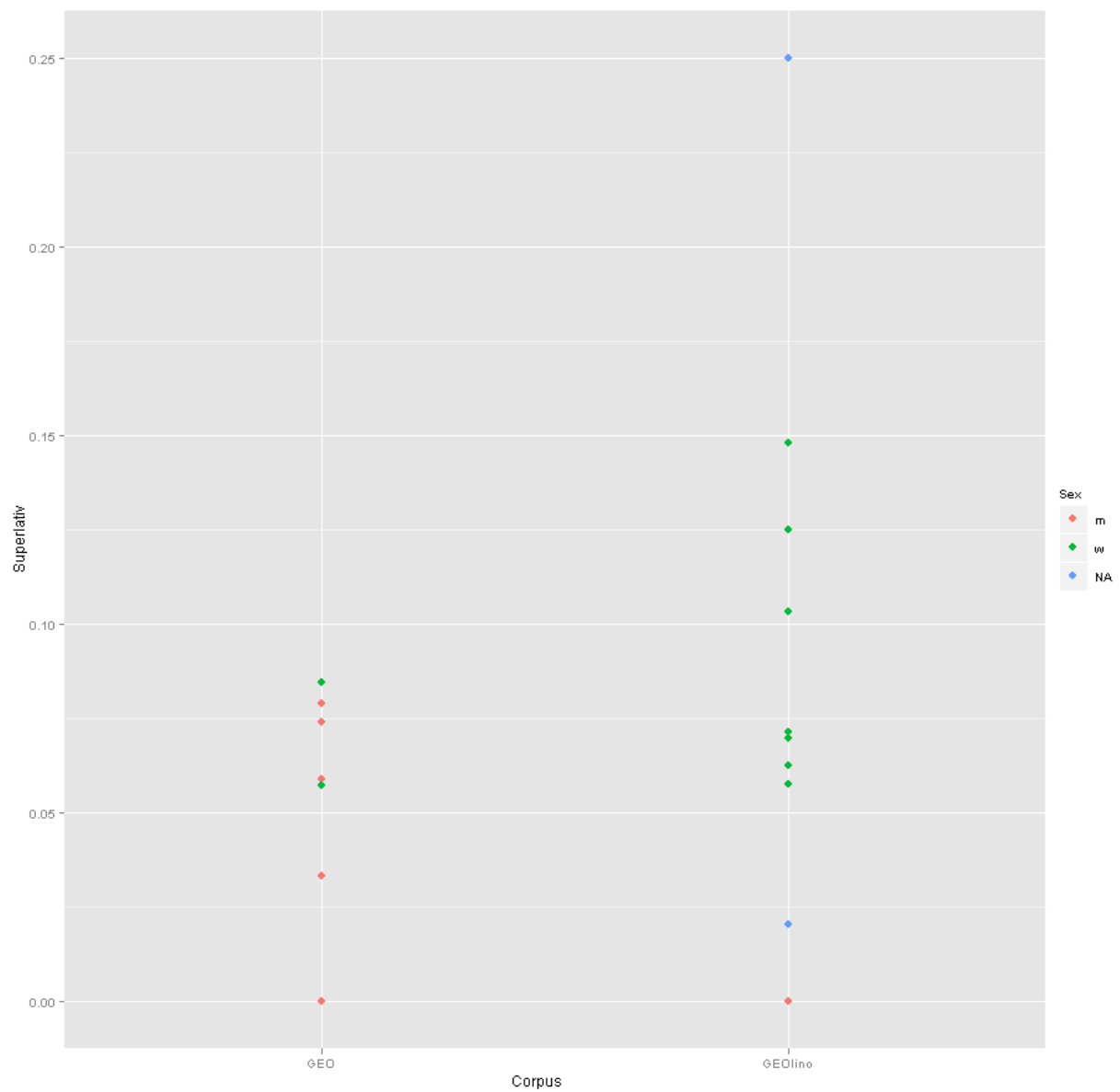


Abbildung 12: Verhältnis von Superlativ zu Positiv und Komparativ

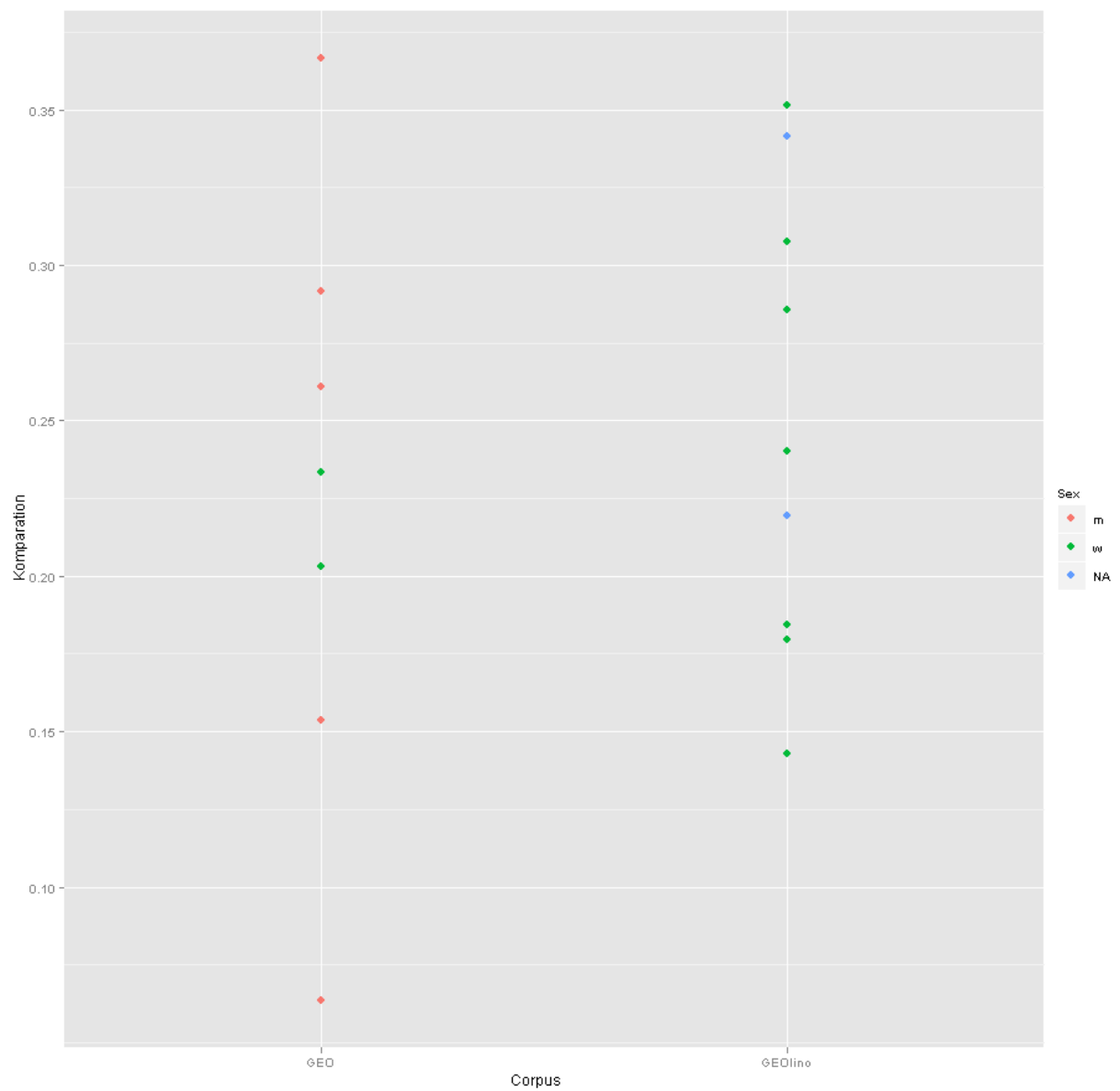


Abbildung 13: Verhältnis von Komparativ und Superlativ zu Positiv

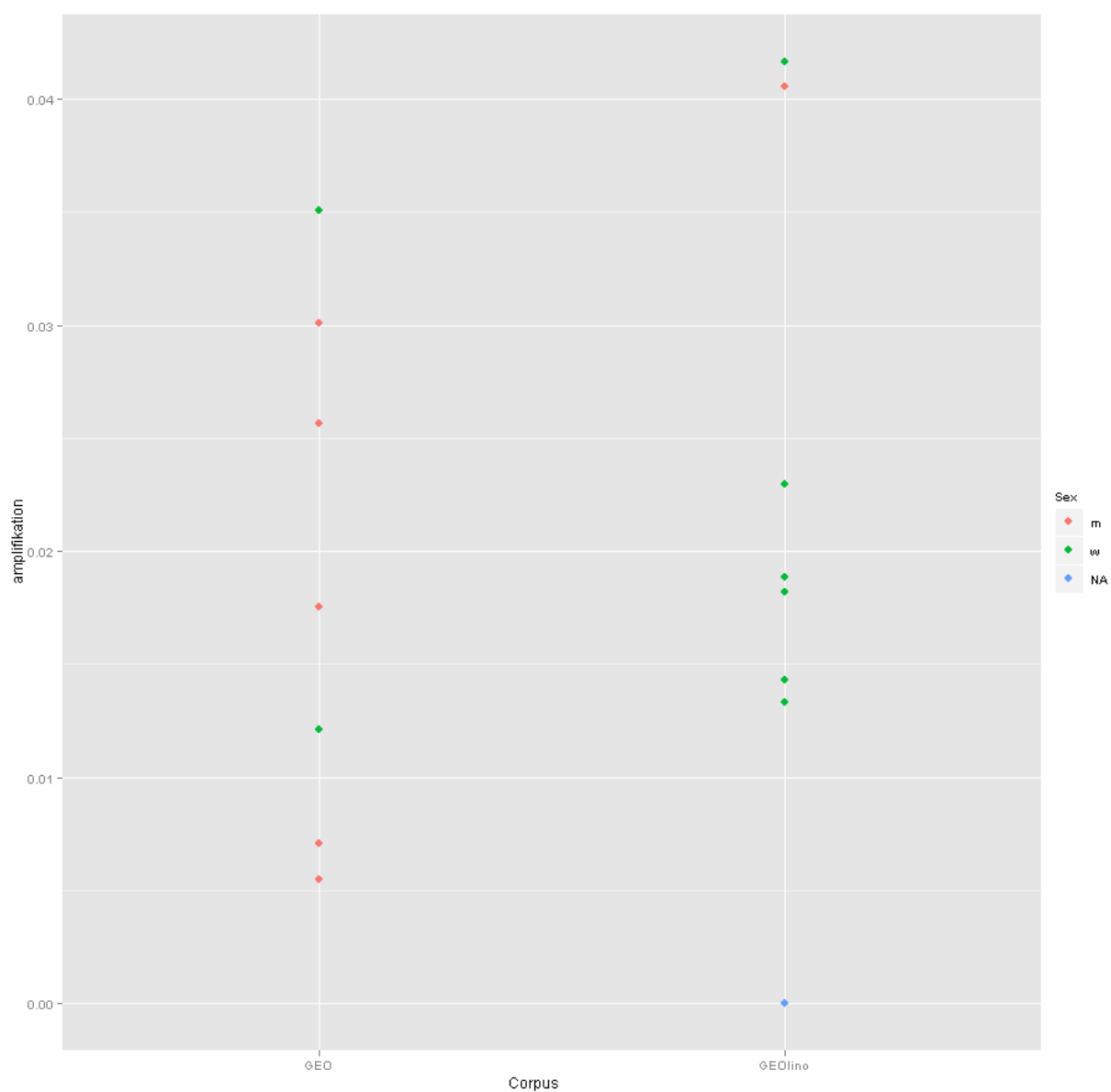


Abbildung 14: Verhältnis amplifizierter Adjektive zu nicht-amplifizierten Adjektiven

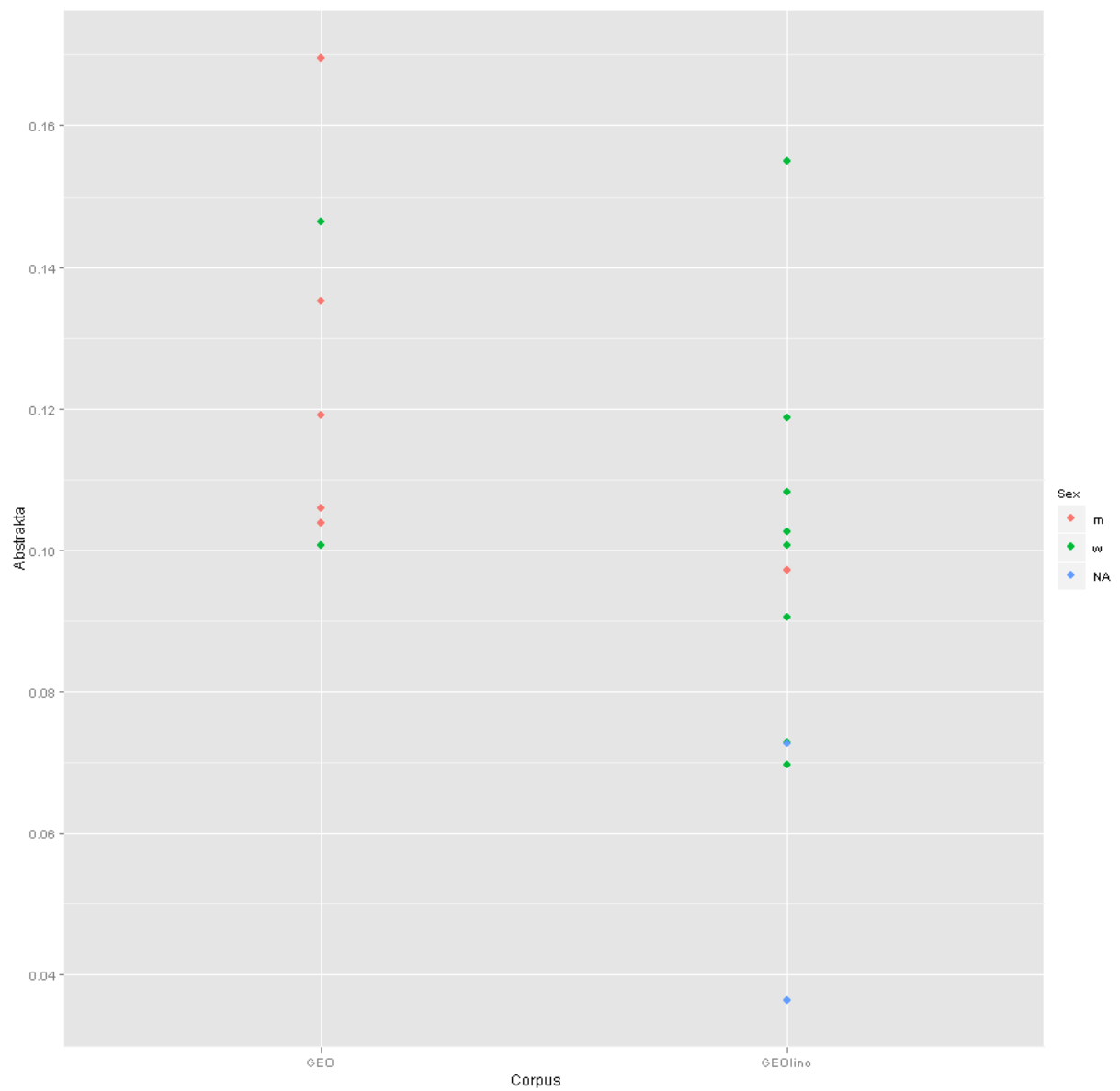


Abbildung 15: Verhältnis von nominalen Abstrakta zu nominalen Konkreta

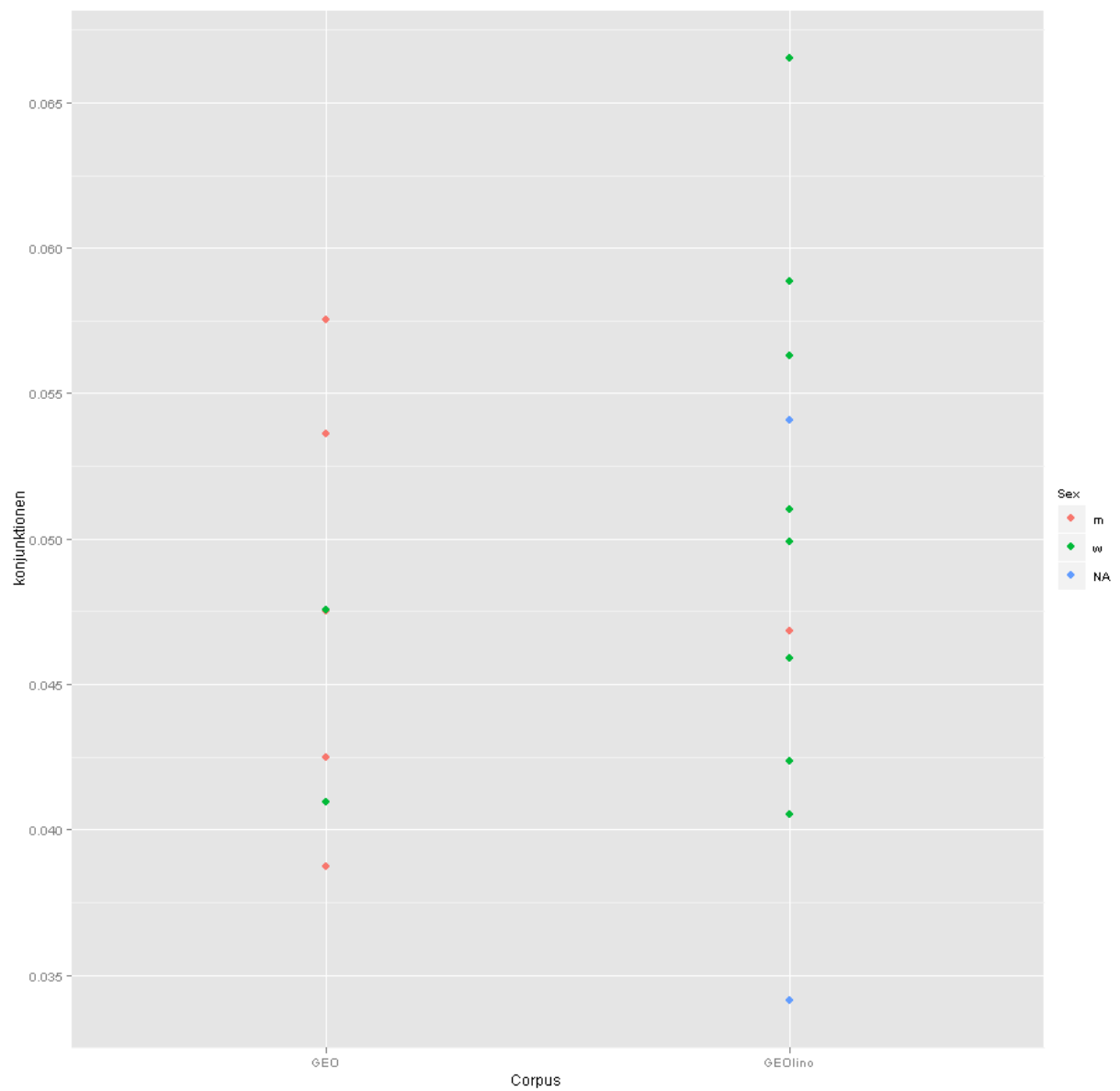


Abbildung 16: Anteil der Konjunktionen an der Gesamt-Tokenzahl

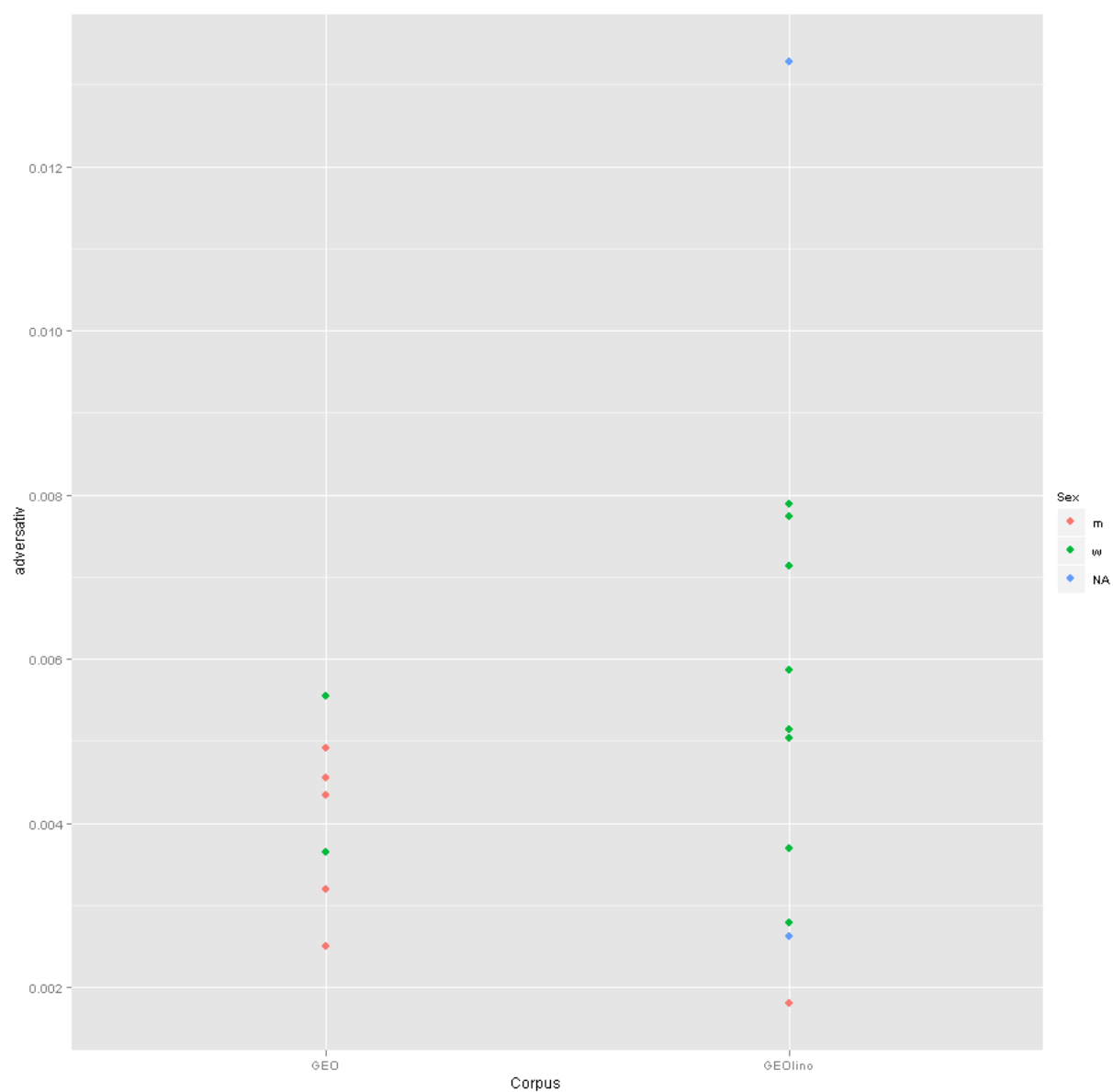


Abbildung 17: Anteil der adversativen Konjunktionen an der Gesamt-Tokenzahl

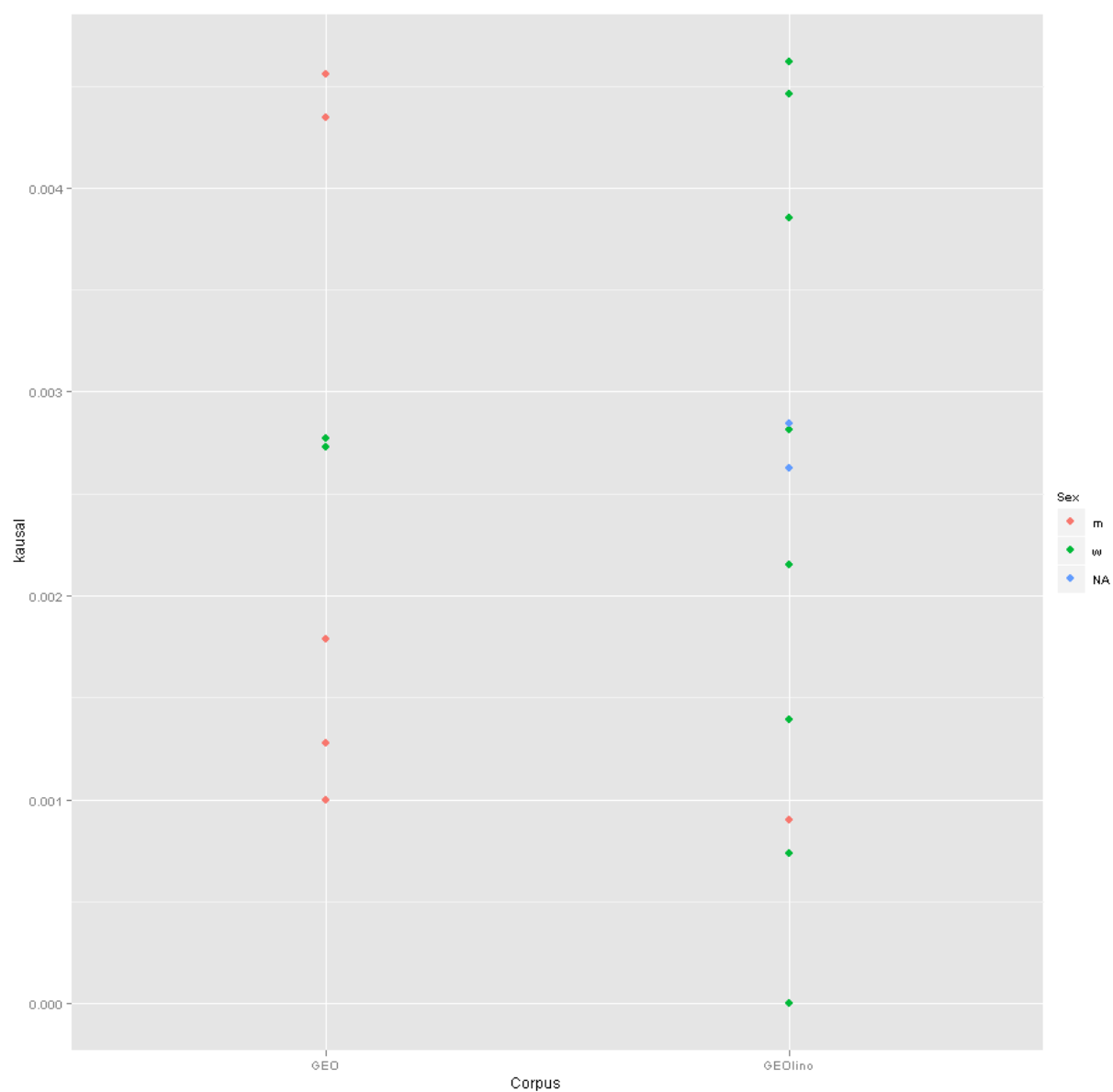


Abbildung 18: Anteil der kausalen Konjunktionen an der Gesamt-Tokenzahl

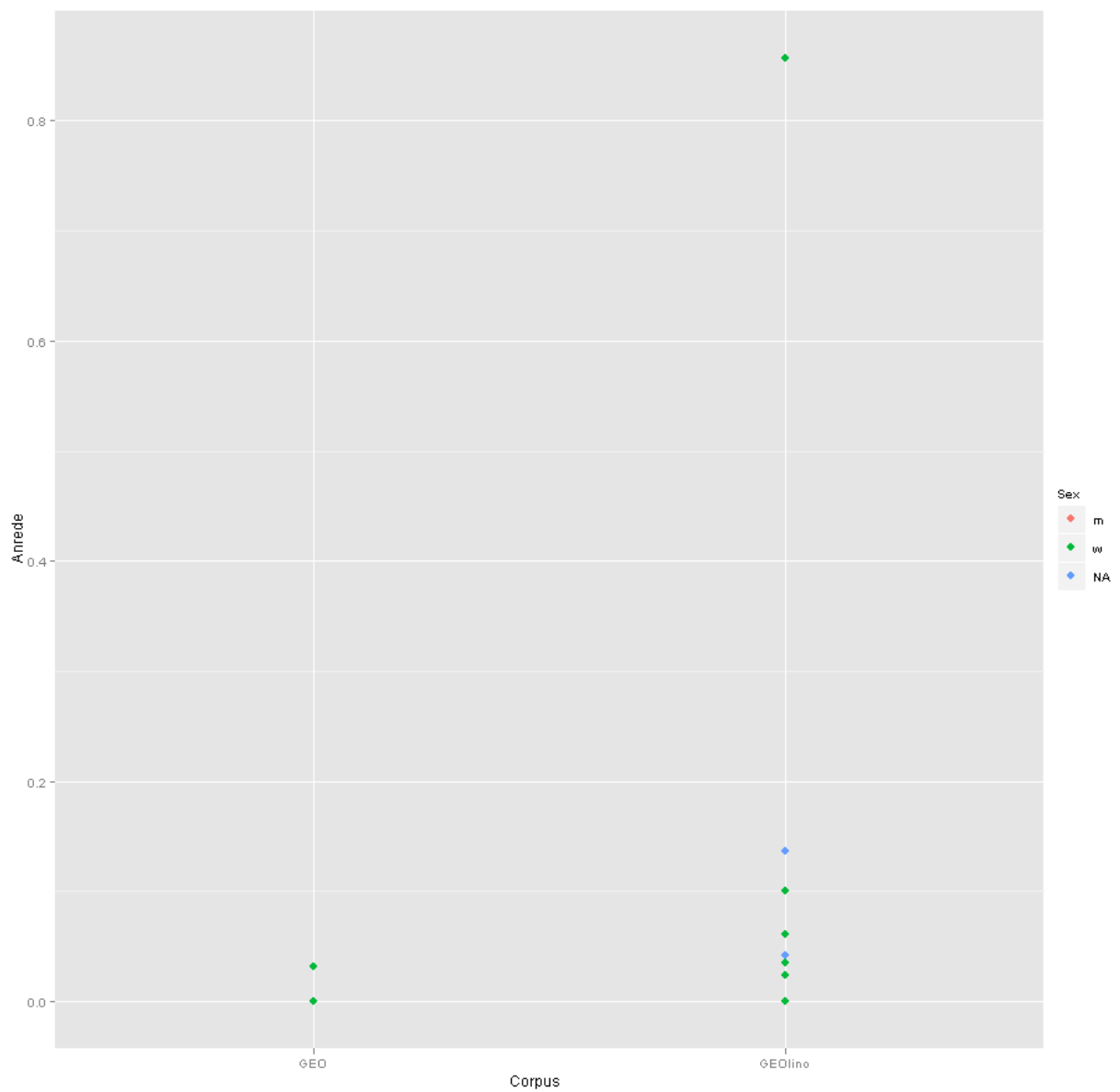


Abbildung 19: Verhältnis von direkter Anrede durch Personalpronomen zu Personalpronomen ohne direkte Anrede

D. Doppeltexte: Verteilung der POS-tags

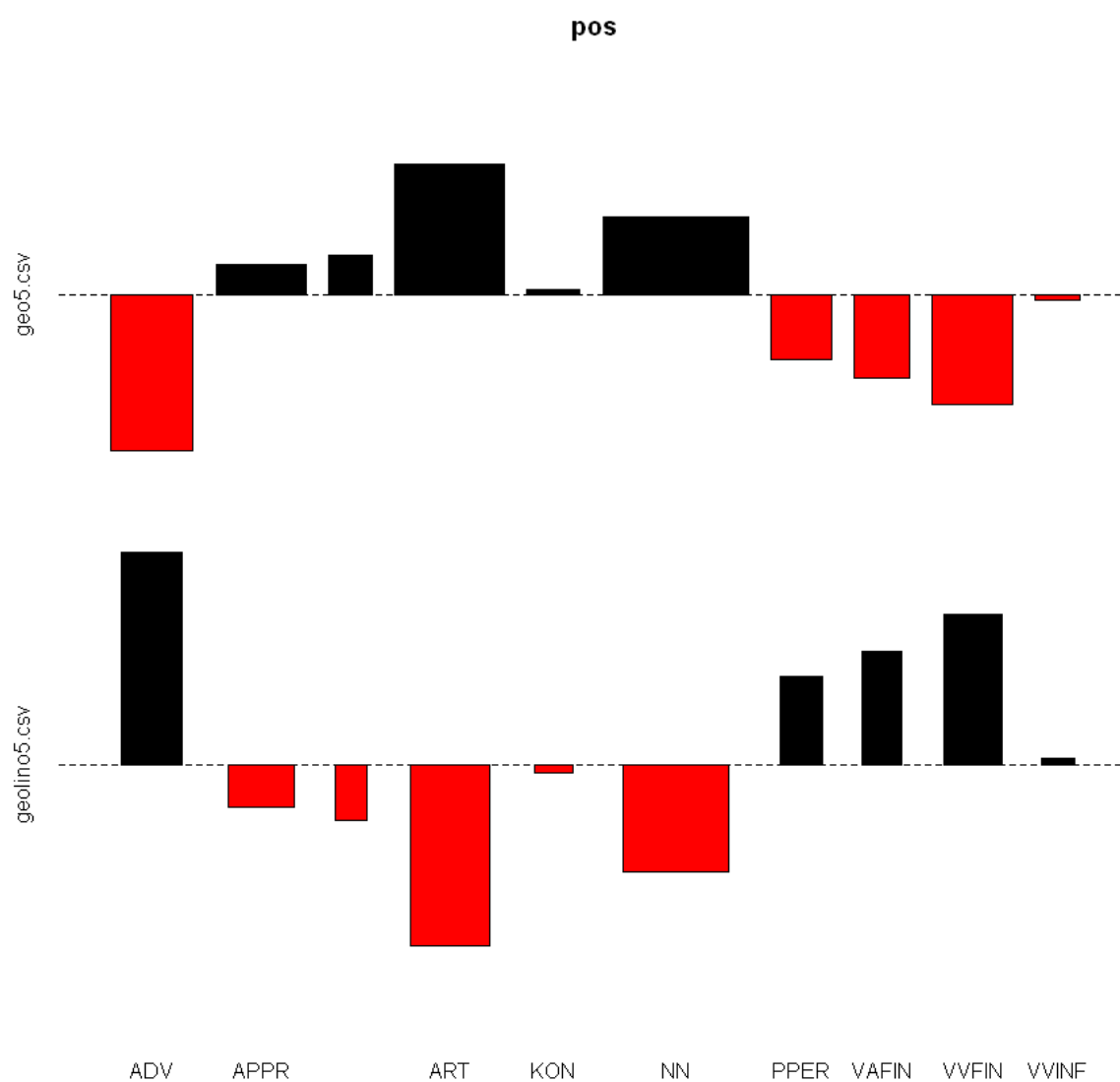


Abbildung 20: Verteilung der POS-tags für *GEOLino5* und *GEO5*

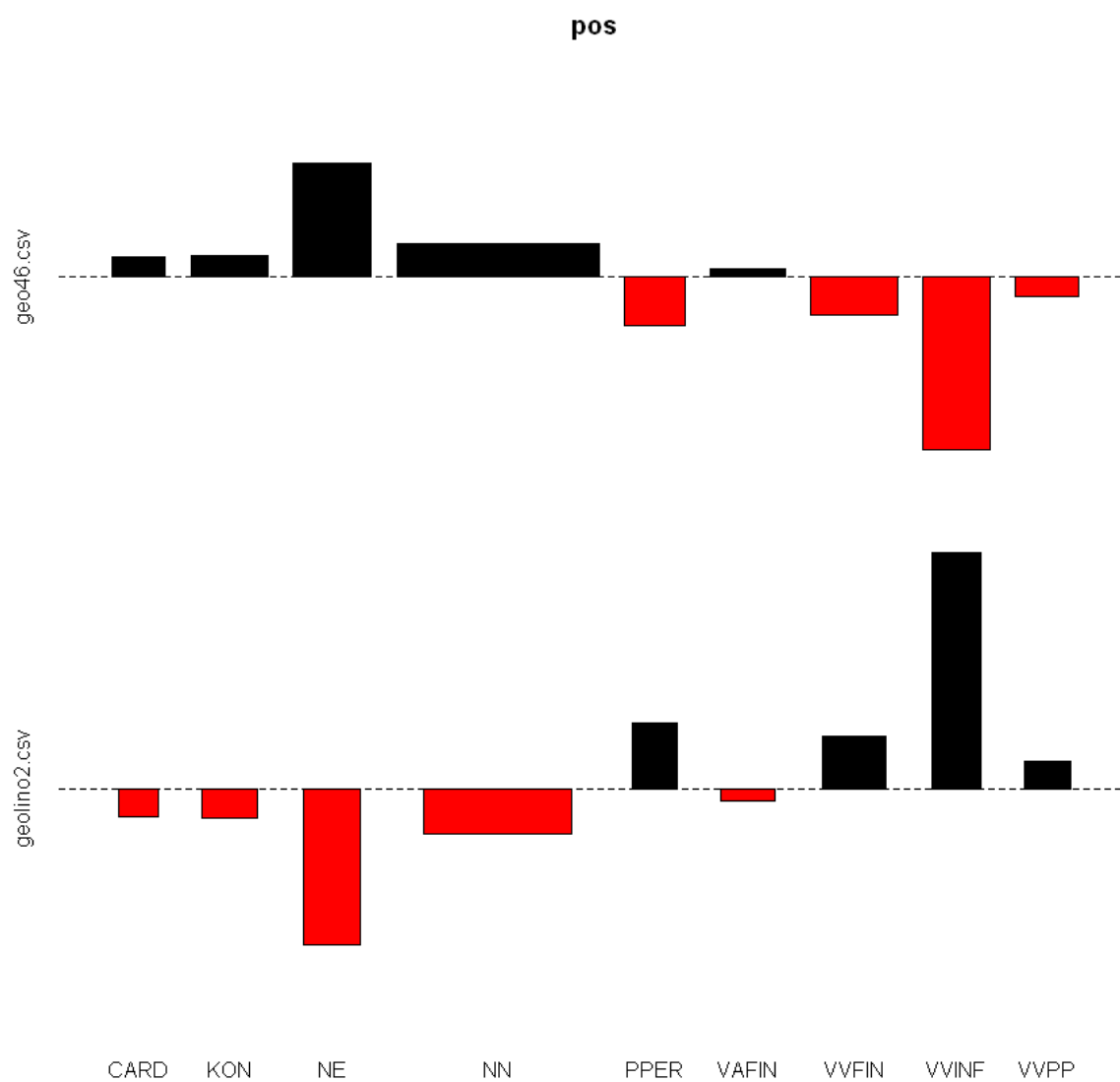


Abbildung 21: Verteilung der POS-tags für *GEOlino2* und *GEO46*

9. Selbstständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit mit dem Titel

Die Sprache in Kinderzeitschriften

Ein korpusbasierter Vergleich der Sprache in *GEOlino* und *GEO*

selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen.

Berlin, den 24. April 2012

Nadine Lestmann

Danksagung

Ich bedanke mich bei Thomas Krause und Carolin Odebrecht für die Anleitung zur Korpuserstellung.

Ich danke auch Jasmine Bennöhr, die mir im Vorfeld meiner Arbeit freundlicherweise ihre Scripte zur Ermittlung der Wort- und Satzlängen zur Verfügung gestellt hat.

Außerdem möchte ich Julia Richling ganz herzlich danken für die Erstellung der Hauptkomponentenanalyse.

Ich danke besonders Felix Golcher, der mich bei der Arbeit in *R* vom Einlesen der Daten bis hin zur Analyse die ganze Zeit unterstützt hat.